Making Effective Maps: Cartographic Design for GIS

By: Sarah Schlosser, GISP

University of Colorado Boulder
Department of Geography

The creation of this work *Making Effective Maps: Cartographic Design for GIS* was supported by Open CU Boulder 2023-2024, a grant funded by the Colorado Department of Higher Education with additional support from the CU Office of the President, CU Office of Academic Affairs, CU Boulder Office of the Provost, and CU Boulder University Libraries.

TABLE OF CONTENTS

Chapter 1
Introduction to Cartography

1.0: Defining a Map

Maps are a graphic representation of the earth, showing the spatial relationship of the earth's features whether cultural or physical. Central to this representation is reducing the earth's features of interest to a manageable size (i.e., map scale) and its transformation into a functional two-dimensional form (i.e., map projection). As all maps are visual representations of selected phenomena, a cartographer must consider the key concepts that make a map a map: it is a scaled, selective, symbolized, and abstracted graphical representation of the phenomenon [4] [5].

Scaled: Scaled refers to the fact that almost all maps are more useful when they are reproduced at a smaller scale than the actual phenomenon being mapped. For example, a road map of Idaho that when unfolded is the actual size of Idaho, is not very useful to us, so instead, we reduce the size of the spatial phenomenon being represented onto a more reasonably sized piece of paper [5].

Selective: Selective means that maps should only include items that are directly related to the message of the map. For instance, a map designated to show where all the hospitals are located within a city should not include the location of every manhole cover in the city. The reason being, the location of the manhole covers is irrelevant to the locations of the hospitals, and therefore it would be a meaningless feature on the map and likely detrimental to achieving the map's purpose [5].

Symbolized: Symbolized, refers to the idea of extracting the item being mapped by using a representative symbol. Examples of representative symbols would be a star with a circle around it that represents a state capital, or a symbol of the front of a bus that represents a bus stop [5].

Abstracted: Decisions about how to classify, simplify, or exaggerate features and how to symbolize objects of interest simultaneously fall under the realms of art and science. Moving from the real world to the world of maps is map abstraction. This process involves making choices about how to represent features. Regarding geographic information systems (GIS), we must be explicit, consistent, and precise in defining and describing geographical features of interest [4]

1.1 Defining Cartography

Cartography is the art and science of making maps along with their study as scientific works of art. It involves the process of producing a map through the philosophical and theoretical basis of map-making. A *cartographer* is someone who designs and prepares a map for distribution. More specifically, a cartographer is someone who studies the philosophical and theoretical basis of the rules for making maps. Traditionally, only cartographers made maps and these cartographers were considered skilled workers that required many years of apprenticeship and practice. However, in recent history, cartography has been democratized and now many different groups create maps without the use of someone specifically trained in cartography [5].

Cartography is seldom a stand-alone profession and is now seen as a skill set possessed by geographers, geographic information scientists, or anyone else who wishes to create a map. The profession of cartography still does exist and is important to many fields of study including environmental science, business, and epidemiology [5].

1.2 Defining GIS

Geographic information systems (GIS), are computer systems used to store, visualize, analyze, manage, edit, and interpret geographic data. GIS data, often referred to as spatial or geospatial data, includes both attribute and spatial information about a feature. For example, you may have data for Boulder County that indicates both its geospatial location on the earth's surface and have attributes about the county in tabular format such as the number of residents, percentage of people living in poverty, and median household income. The first GIS was developed in the early 1960s by Roger Tomlinson, referred to as the "father of GIS" [6]. For more information about the history of GIS and spatial analysis see "A Brief History of GIS" https://www.gislounge.com/history-of-gis/ [opens in new tab].

What is GIS? Video https://youtu.be/LHDCRjAxpI0?si=kUC3VFVSV49vztIp [opens in new tab]

1.3 Types of GIS Data [2]

Within the realm of maps, cartography, and GIS, the world is made up of various features. Such features include but are not restricted to points of interest, roads, rivers, lakes, soils, oceans, and buildings. Moreover, such features have a form, and more precisely, a geometric form. For instance, cities and power poles are considered point-like features; rivers and streams are linear features; and lakes, countries, and forests are areal (aka polygon) features.

Features can also be categorized as either discrete or continuous. Discrete features are well-defined and are easy to locate, measure, and count, and their edges or boundaries are readily defined. Examples of discrete features include buildings, roads, and parks. Continuous features are less well-defined and exist across space. The most commonly cited examples of continuous features are temperature and elevation. Changes in both temperature and elevation tend to be gradual over relatively large areas.

Geographical features also have several characteristics, called attributes, associated with the spatial feature. For instance, a forest may be further described by its attributes, for instance, whether it is made up of deciduous, coniferous, or a mix of both types of trees. More general attributes may include measurements such as tree density per acre or average canopy height in meters.

GIS data must include the spatial location and extent of the feature(s), which allows us to define its location on the surface of the earth. Such data include but are not limited to the latitude and longitude coordinates of points of interest, street addresses, postal codes, political boundaries, and even the names of places of interest. It is also important to note the difference between spatial data and attribute data, whereas spatial data are concerned with defining the location of an object of interest, attribute data are concerned with its nongeographic traits and characteristics.

1.4 GIS Files Formats

Geospatial data are stored in many different file formats. Each geographic information system (GIS) software package and version supports different formats. Although several of the more common file formats are summarized here, many other formats exist for use in various GIS programs [2].

Vector File Formats: The most common vector file format is the shapefile. Shapefiles, developed by ESRI in the early 1990s, are simple, files developed to store the geometric location and attribute information of geographic features. Shapefiles are incapable of storing null values. Field names within the attribute table are limited to ten characters, and each shapefile can represent only one type of geometry, i.e. either point, line, or polygon features. Supported data types are limited to floating point, integer, date, and text. Shapefiles are supported by almost all commercial and open-source GIS software [2].

Despite being called a “shapefile,” this format is actually a compilation of many different files. Figure 1.0 lists and describes the different file formats that can associated with a shapefile. Among those listed, only the SHP, SHX, and DBF file formats are mandatory to create a functioning shapefile [2].

File Extension	Purpose
SHP*	Feature geometry
SHX*	Index format for the feature geometry
DBF*	Feature attribute information in dBASE IV format
PRJ	Projection information
SBN and SBX	Spatial index of the features
FBN and FBX	Read-only spatial index of the features
AIN and AIH	Attribute information for active fields in the table
IXS	Geocoding index for read-write shapefiles
MXS	Geocoding index for read-write shapefiles with ODB format
ATX	Attribute index used in ArcGIS 8 and later
SHP.XML	Metadata in XML format
CPG	Code page specifications for identifying character encoding
*** Indicates mandatory files**	

Figure 1.0. Shapefile extensions and purposes.
Credit: Shapefile file types, *Essentials of GIS*, Campbell and Shin, CC BY-NC-SA 3.0

Raster File Formats: A multitude of raster file format types are available for use in GIS. Due to ongoing technological advancements, raster image file sizes have increased significantly. To deal with this potential constraint, two types of file compression are commonly used: lossless and lossy. Lossless compression reduces file size without decreasing image quality. Lossy compression attempts to exploit the limitations of the human eye by removing information from the image that cannot be sensed. Lossy compression results in smaller file sizes than lossless compression [2].

Among the most common raster files used on the web are the JPEG (Joint Photographic Experts Group), TIFF (Tagged Image File Format), and PNG (Portable Network Graphics) formats, all of which are open source and can be used with most GIS software packages. However, native JPEG, TIFF, and PNG files do not have georeferenced information associated with them and therefore cannot be used in any geospatial mapping efforts without them first being georeferenced [2], meaning that a geographic coordinate system has been assigned to the data. Some raster data is available in GeoTIFF format which has already been georeferenced.

An example of a raster file format with explicit georeferencing information is the proprietary MrSID (Multiresolution Seamless Image Database) format. This lossless compression format was developed for use with large aerial photographs or satellite images [2].

Some raster file formats are developed explicitly for modeling elevation. The USGS DEM (US Geological Survey Digital Elevation Model) is a popular file format due to widespread availability and extensive software support for the format. Each pixel value in these grid-based DEMs denotes spot elevations on the ground, usually in feet or meters. DEMs are referred to as digital terrain models

(DTMs) when they represent a bare-earth model and as digital surface models (DSMs) when they include the heights of landscape features such as buildings and trees [2].

Hybrid File Formats: Geodatabases have been developed in a proprietary ESRI file format that supports both vector and raster feature datasets (e.g., points, lines, polygons, annotation, JPEG, TIFF) within a single file, which serves as a container for multiple data layers and information [2]. Geodatabases function as containers for organizing and storing geospatial data, with different types of geodatabases having different functionality and size limits.

There are three different types of geodatabases. The **personal geodatabase** was developed for single-user editing, whereby two editors cannot work on the same geodatabase at a given time. The personal geodatabase employs the Microsoft Access DBMS file format and maintains a size limit of 2 gigabytes per file. The personal geodatabase is used by ESRI ArcGIS versions 8.x to 10.x; it is not compatible with any versions of ArcGIS Pro [2].

The **file geodatabase** similarly allows only single-user editing, but this restriction applies only to unique feature datasets within a geodatabase. The file geodatabase incorporates new tools such as domains (rules applied to attributes), subtypes (groups of objects with a feature class or table), and split/merge policies (rules to control and define the output of split and merge operations). This format has a size limit of 1 terabyte. File databases are not tied to any specific relational database management system (e.g., Microsoft Access) and can be employed on both Windows and UNIX platforms [2].

The **ArcSDE geodatabase** allows multiple editors to simultaneously work on feature datasets within a single geodatabase (a.k.a. versioning). Like the file geodatabase, this format can be employed on both Windows and UNIX platforms. File size is limited to 4 gigabytes and its proprietary nature requires an ArcInfo or ArcEditor license for use. The ArcSDE geodatabase is implemented on the SQL Server Express software package, which is a free DBMS platform developed by Microsoft [2].

References - materials are adapted from the following sources:

[1] Cartography Guide by Axis Maps under a CC BY-NC-SA 4.0 license

[2] Essentials of Geographic Information Systems by Saylor Academy under a CC BY-NC-SA 3.0 license

[3] GEOG 3053 Cartographic Visualization by Barbara Buttenfield, University of Colorado Boulder, used with permission.

[4] Geographic Information Systems and Cartography by Adam Dastrup under a CC BY-NC-SA 4.0 license

[5] Introduction to Cartography by Ulrike Ingram under a CC BY 4.0 license

[6] Introduction to GIS by Victor Olaya under a CC BY 3.0 license

Chapter 2
Types of Maps

2.0 Map Mediums

Classifying maps begins with categorizing them into mediums. The three types of map mediums are tangible, virtual, and mental [5].

Tangible Map: A tangible map is a map that you can hold in your hands such as a paper map. Tangible maps are portable and can be stored for long amounts of time. Additionally, no specialized software, hardware, or even internet access is required to access these maps. However, the disadvantage is that they only represent one point in time, therefore their currentness may be in doubt [5].

Virtual Map: A virtual map is any map displayed on a computing device. Maps can be viewed on desktop computers, tablets, laptops, phones, GPS receivers, and many other digital devices. Virtual maps are easily updatable, can be dynamic, can show animation, can link to large amounts of information such as documents, pictures, movies, and sounds, and can be easily shared. Negative aspects of virtual maps include that they require hardware and software to view, require maintenance, may not be intuitive to many users, and may require more expertise to create [5].

Mental Map: A mental map (Figure 2.0) is a type of virtual map that is stored in one's mind and is their conceptualization of space. Mental maps do not translate exactly from person to person except through the conversion of the mental map to a tangible or virtual map, or to any other communication path such as speech or writing [5]. We rely on our mental maps to get from one place to another, to plan our daily activities, or to understand and situate events that we hear about from our friends, family, or the news. Mental maps also reflect the amount and extent of geographic knowledge and spatial awareness that we possess [2].

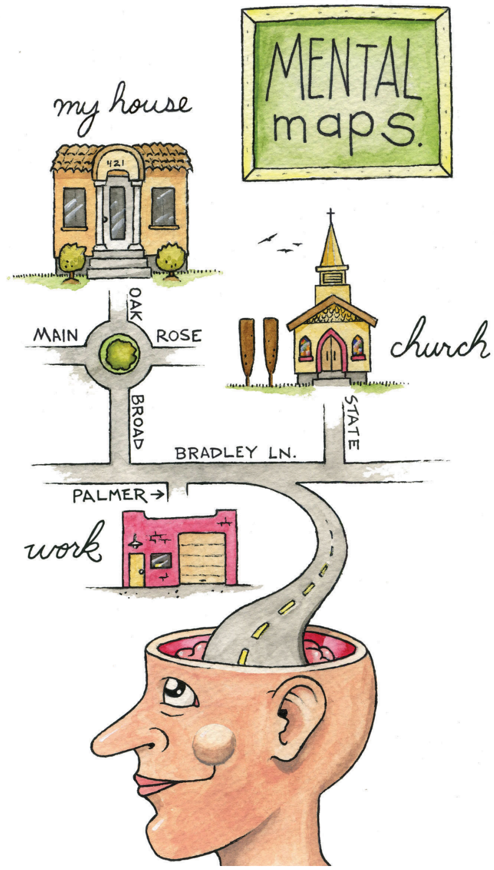


Figure 2.0. Mental map example
Credit: Image from *Essentials of GIS*, Campbell and Shin, CC BY-NC-SA 3.0

2.1 Types of Maps

Reference Map: A general reference map emphasizes location and shows a variety of features. A general reference map primarily displays objects, their location, and identifying information via labels. They are typically not specialized to any one particular use but are instead meant to be used for a wide variety of activities and purposes [5]. A reference map's geographic features and map elements tend to be treated and represented equally. In other words, no single aspect of a reference map takes precedence over any other aspect. Moreover, reference maps represent geographic reality accurately. Examples of standard reference maps include topographic maps created by the United States Geological Survey (USGS) and image maps obtained from satellites or aircraft available through online mapping services [4].

Thematic Map: A thematic map emphasizes attributes related to a subject(s) or theme(s). One of the reasons thematic maps are so useful is that they display patterns of the theme across space. Thematic maps focus on a theme(s) which is commonly referred to as an attribute(s). While most thematic maps focus on a single theme it is important to note that thematic maps can also display multiple related attributes at the same time - these maps are referred to as multivariate thematic maps [2].

If multiple thematic maps are created relating to a single, or multiple related themes, it can be very useful to compare the two maps to draw additional conclusions. Unlike general reference maps, thematic maps are typically very selective in the features included on the map. Only features that support the map's theme are to be included on thematic maps. For example, county boundaries are indicated on a thematic map of election results, but highways and rivers are not included since they are not relevant to the theme being mapped [2].

A special purpose map called a **pragmatic map**, is designed specifically to guide spatial behavior such as how to get from one location to another. The London Underground Map is an example of this type of map. It is designed to allow the user to choose the appropriate rail line and station with ease, excluding unnecessary surface information and generalizing the rail line paths. Other examples of pragmatic maps include aeronautical and nautical charts which are used for traveling over land or sea [3].

Propaganda / Persuasive Map: A propaganda map, more recently referred to as a persuasive map, is designed to influence the reader's beliefs and opinions about a particular subject or event (e.g., war) often imparting fear or offense through the use of exaggerated figures such as octopus or soldiers. The Cornell University Library, Division of Rare & Manuscript Collections has an online catalog of more than 800 persuasive maps https://persuasivemaps.library.cornell.edu/ [opens in new tab].

2.2 Categories of Thematic Maps

Thematic maps can be broken down into specific categories. The six common categories of thematic maps are choropleth, dot density, proportional symbol, graduated symbol, flow maps, and cartograms [5].

Choropleth Maps: On a choropleth map, each enumeration unit, such as a county, is assigned a color that represents either a single value or range of values (data broken into classes) that exist in that enumeration unit. Figure 2.1 is a choropleth map displaying quantitative information that has been aggregated to the county level.

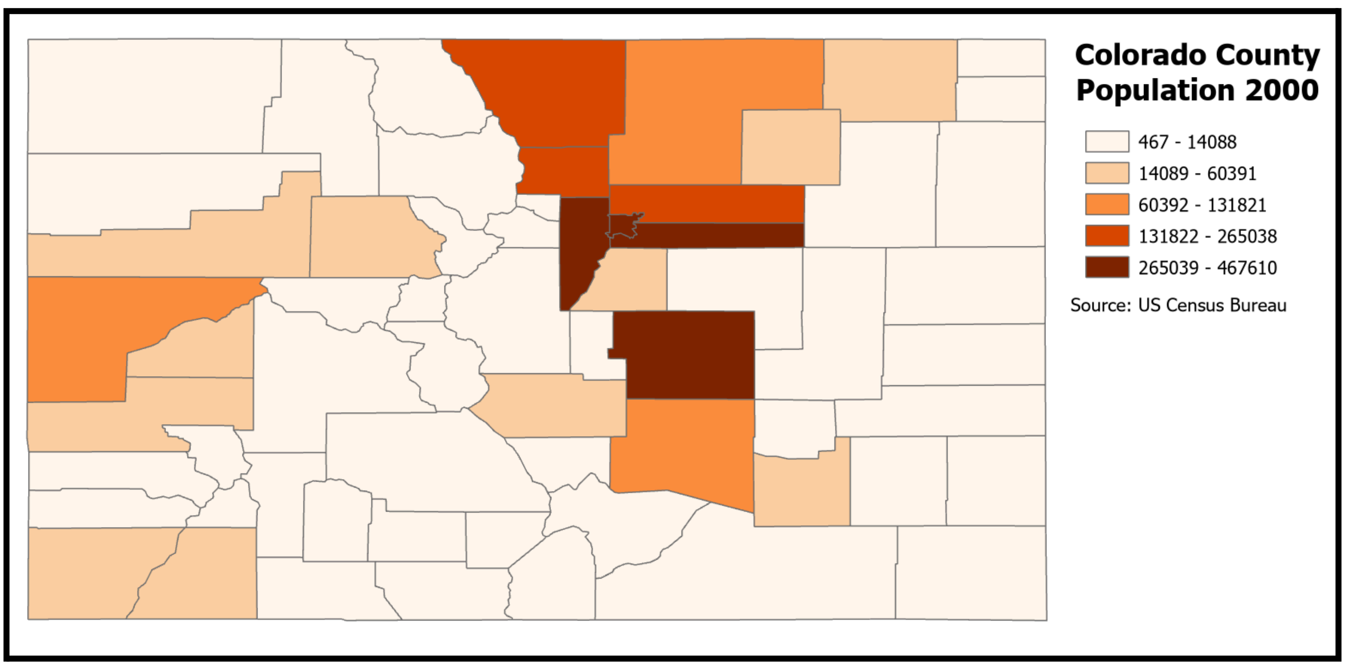


Figure 2.1. Choropleth map.
Data Source: US Census Bureau

Dot Density Maps: A dot density map shows total values of quantitative information represented by dots that are randomly placed within an enumeration unit. Figure 2.2 is a dot density map showing the

Colorado Population where one dot represents 500 people and dots are randomly placed within the counties.

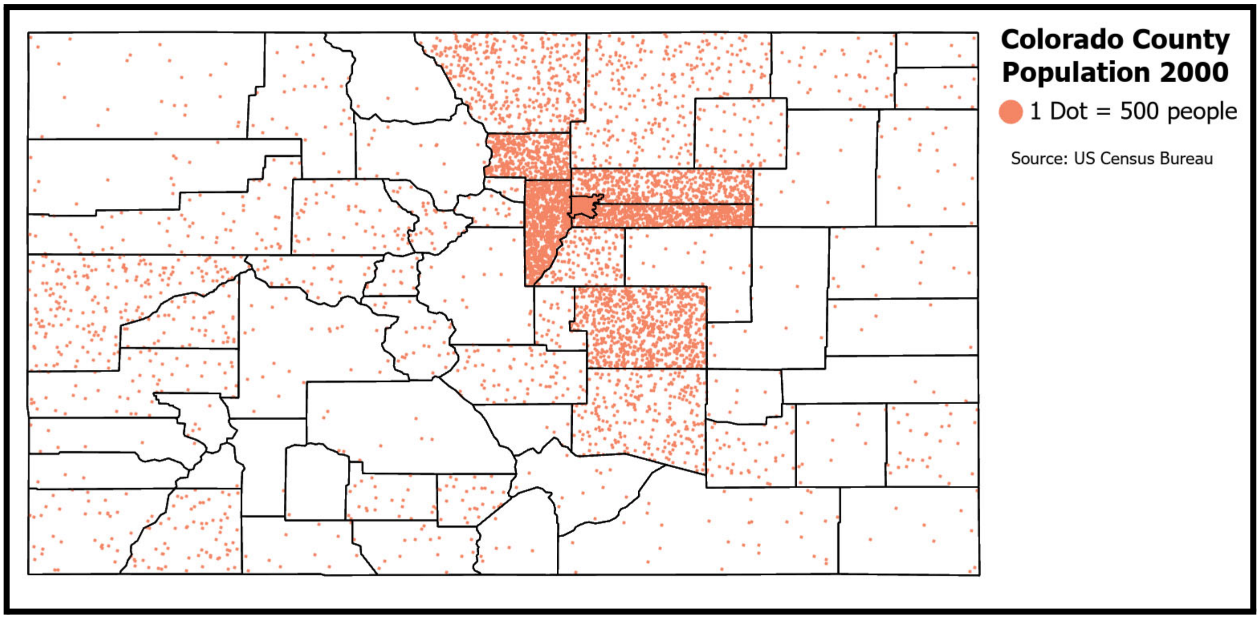


Figure 2.2. Dot density map.
Data Source: US Census Bureau.

Proportional Symbol Maps: In a proportional symbol map, symbols, such as circles, are sized in proportion to the value of an attribute and allow for the comparison of values between enumeration units [5]. For example, a symbol with a value that is twice as large will have a symbol size that is twice as large based on area. Proportional symbols focus on showing differences in magnitudes where each symbol represents a single value. These maps are used to show values at a particular point, such as a city, or a summation over aggregation units such as a county. Figure 2.3 uses proportional symbols to show the Colorado population by county. See Chapter 5 for more information on proportional symbol maps.

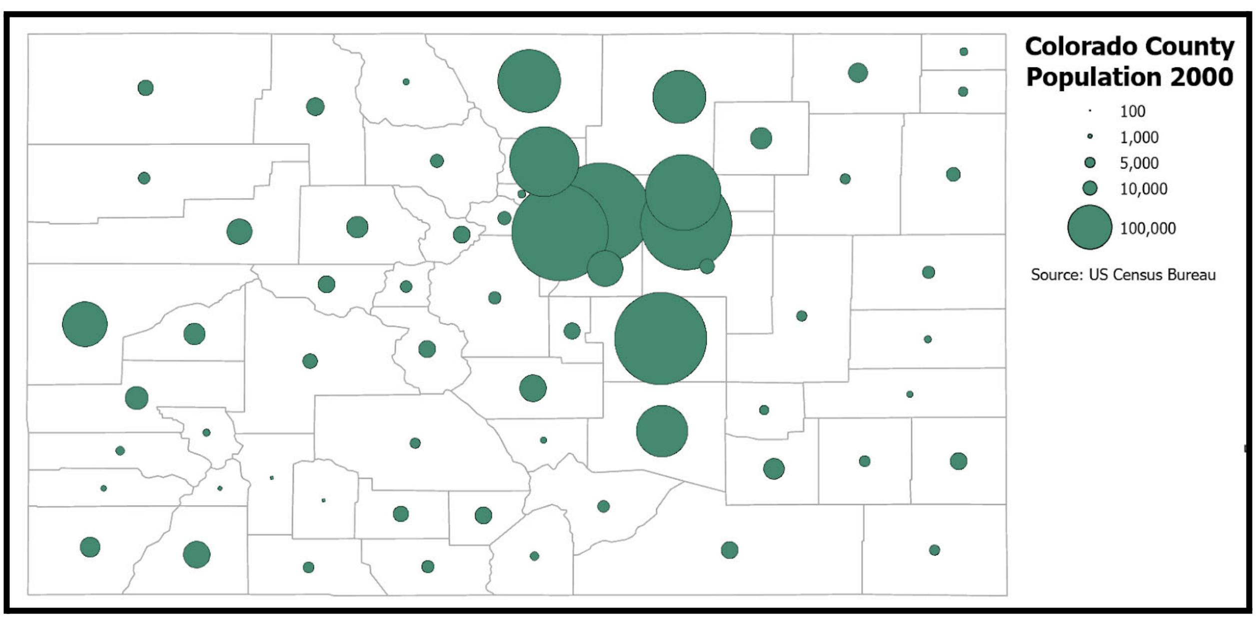


Figure 2.3. Proportional symbol map.
Data Source: US Census Bureau.

Graduated Symbol Maps: Graduated symbol maps are similar to proportional symbol maps as they also use size to show different values at a particular point or sum of values per aggregation units, but graduated symbols represent the order of data values where each symbol represents a range of values. Graduated symbols are not proportional in size. Figure 2.4 uses graduated symbols to show population counts in Colorado. See Chapter 5 for more information on graduated symbol maps.

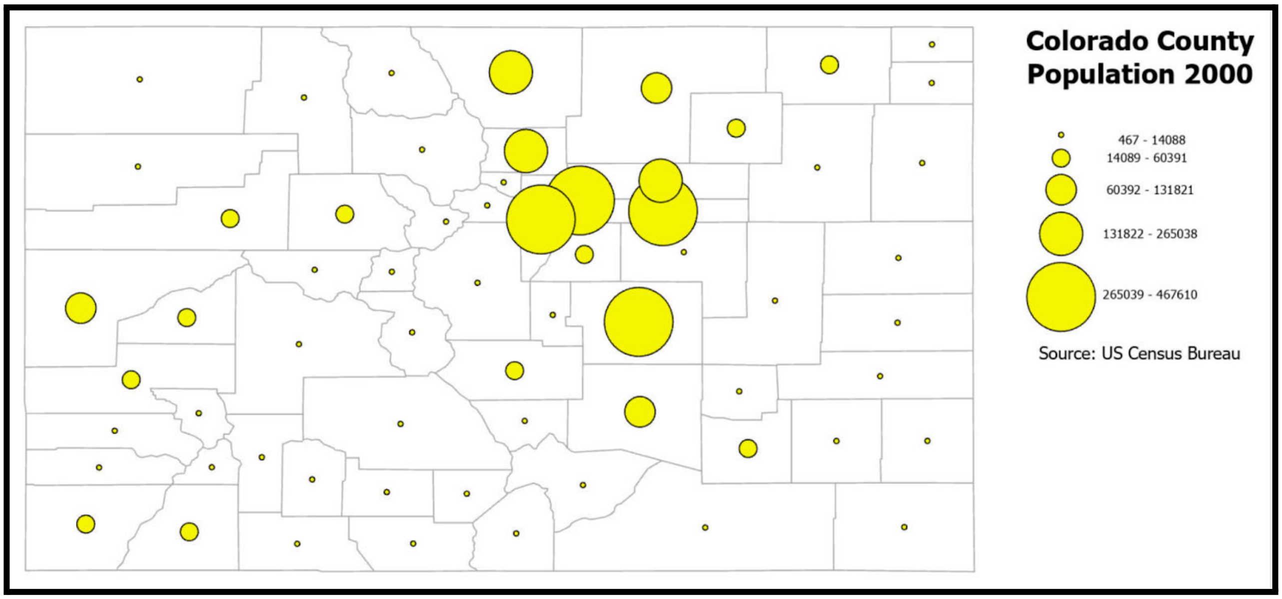


Figure 2.4. Graduated symbol map.
Credit: data from US Census Bureau.

Flow Maps: A flow map uses linear symbols to show the movement of phenomena such as technology, finances, or goods between one location and another (Figure 2.5). Flow maps can use proportional or graduated symbols to show the magnitude or order of values, respectively. These maps often also use color to differentiate flow paths such as for different time periods or different species of animals.

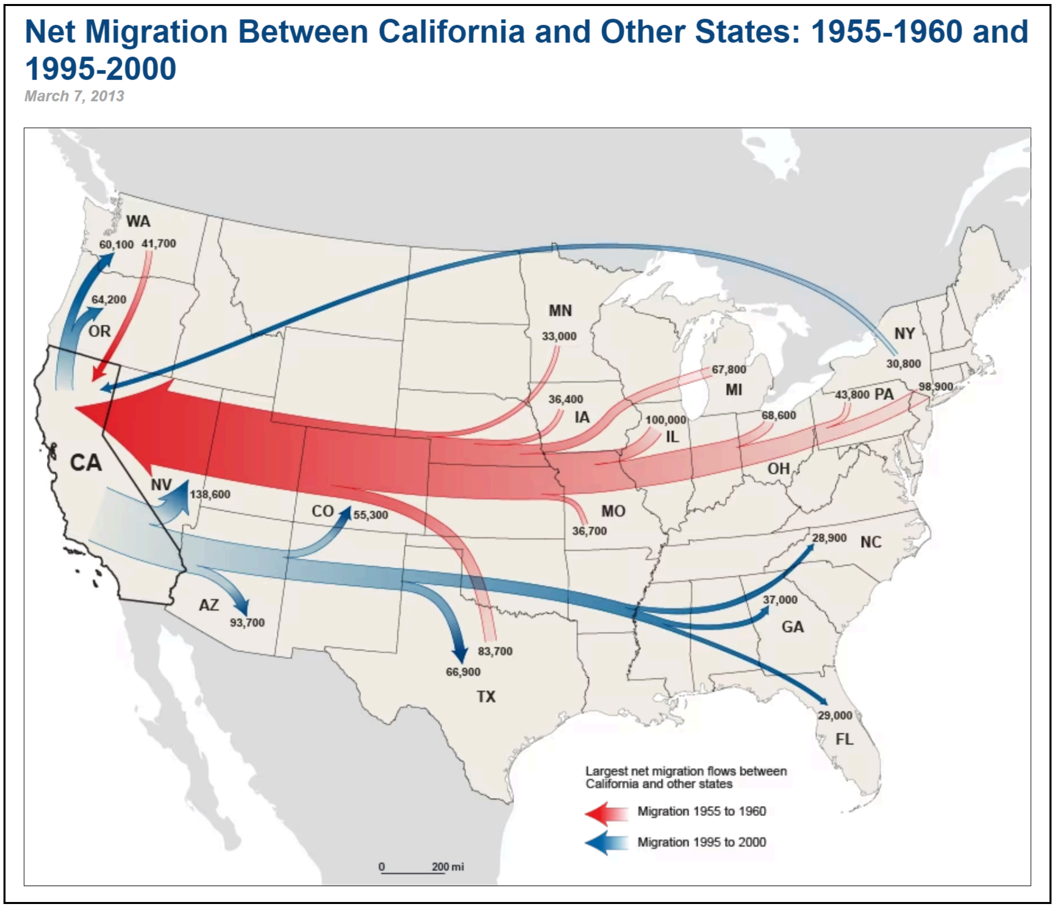


Figure 2.5. Flow map.

Credit: Net Migration Between California and Other States: 1955-1960 and 1995-2000, US Census Bureau, public domain.

Cartograms: In a cartogram map, the area of the enumeration unit, such as a country, is sized in relation to the value of an attribute and allows the map reader to make comparisons relative to the size of each feature. In areas where the value of the attribute is smaller than average, the size of the enumeration unit is reduced, and where the value of the attribute is larger than average, the size of the enumeration unit is enlarged. As the enumeration units are reduced or enlarged relative to the data quantity, the shapes of the areas can become quite distorted (Figure 2.6) [5]. Cartogram maps can be contiguous, as in Figure 2.6, or non-contiguous as in Figure 2.7. Non-continuous cartograms retain the shape of the enumeration unit, therefore are often easier to read especially if basemap units are included as in Figure 2.8. World Mapper https://worldmapper.org/ [opens in new tab] has over 1,100 contiguous cartograms for a variety of themes.

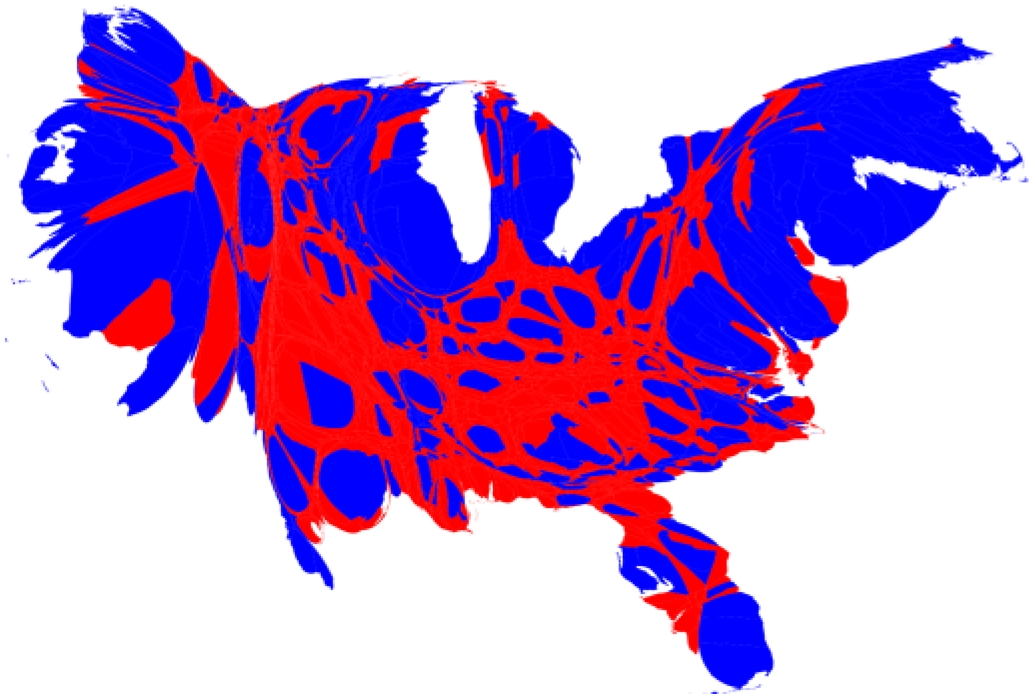


Figure 2.6. Contiguous cartogram showing the 2008 US presidential election results by county.
Credit: from “Maps of the 2008 US presidential election results”, *University of Michigan*, Mark Newman, CC BY 2.0

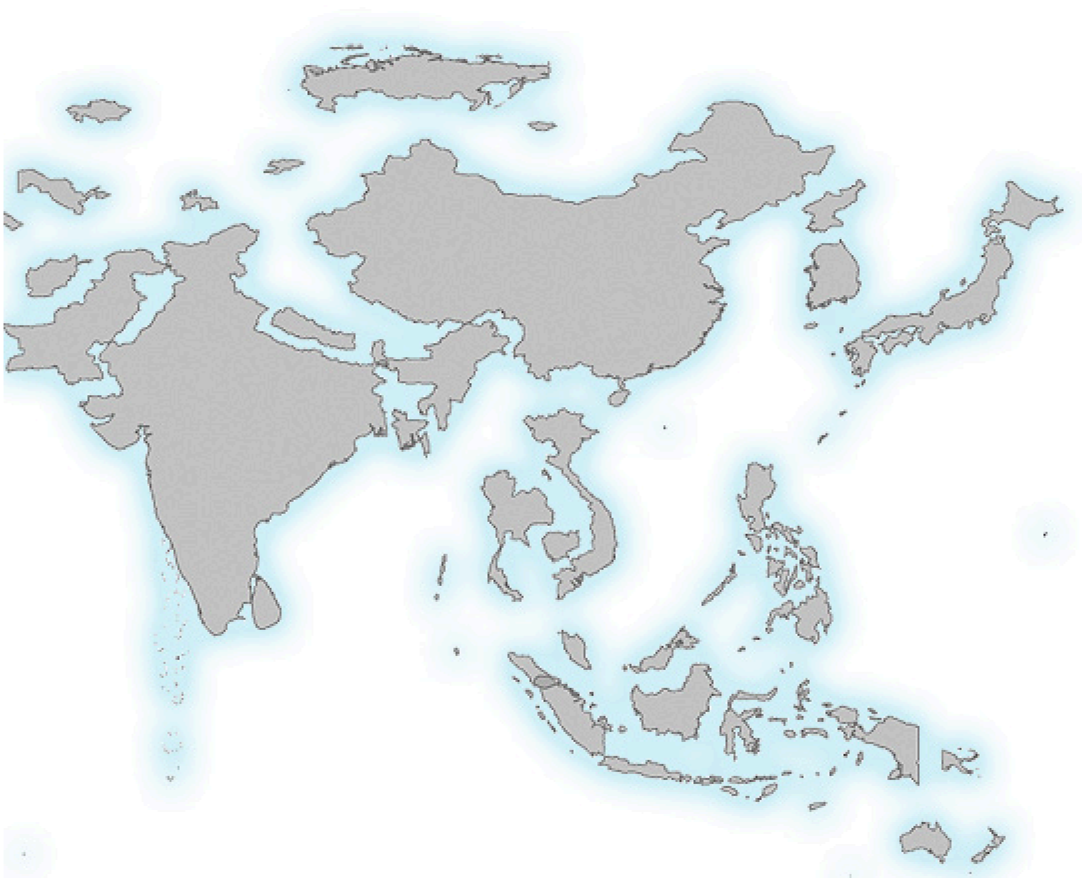


Figure 2.7. Non-contiguous cartogram with countries sized by population.
Credit: Image from *Cartography Guide,* Axis Maps, CC BY-NC-SA 4.0

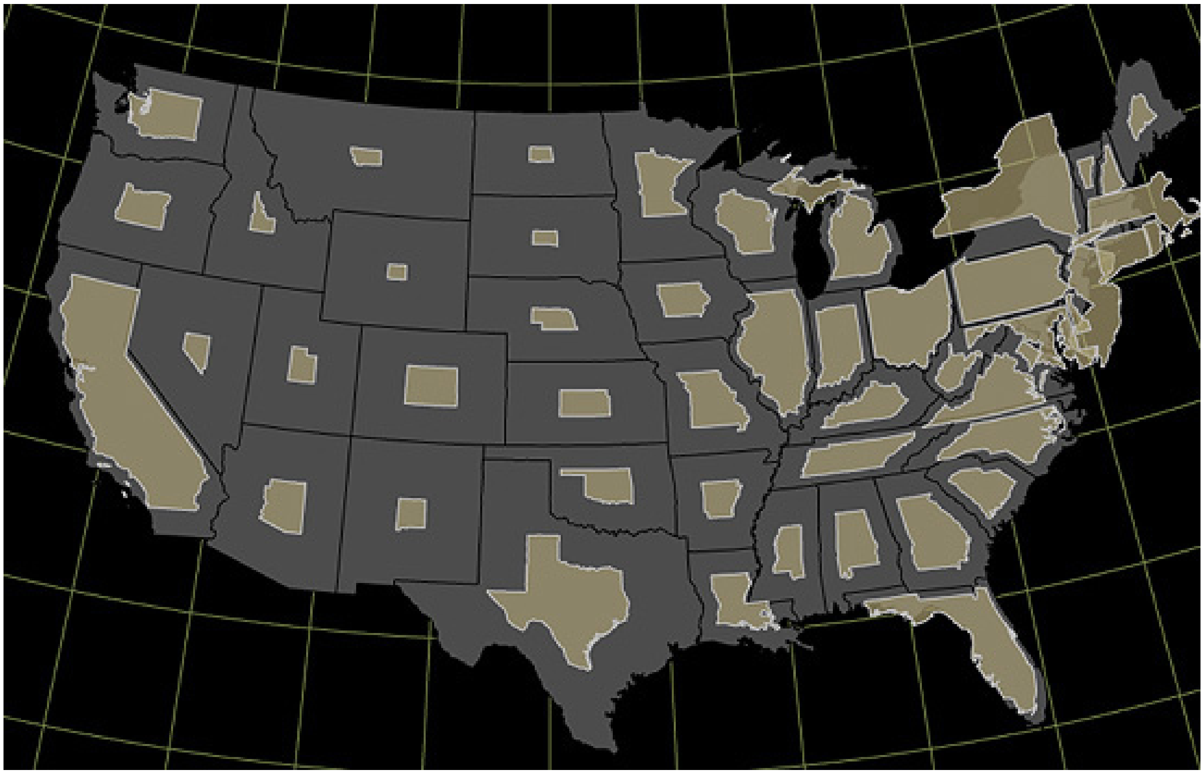


Figure 2.8. Non-contiguous cartogram with enumeration units shown at scale for comparison to cartogram representation for each state.
Credit: Image from *Cartography Guide,* Axis Maps, CC BY-NC-SA 4.0

Additional Resources
Cartograms: https://gistbok.ucgis.org/bok-topics/cartograms [opens in new tab]
Cartography and Art: https://gistbok.ucgis.org/bok-topics/cartography-and-art [opens in new tab]
Cartography and Science: https://gistbok.ucgis.org/bok-topics/cartography-and-science [opens in new tab]

References - materials are adapted from the following sources:

[1] Cartography Guide by Axis Maps under a CC BY-NC-SA 4.0 license

[2] Essentials of Geographic Information Systems by Saylor Academy under a CC BY-NC-SA 3.0 license

[3] GEOG 3053 Cartographic Visualization by Barbara Buttenfield, University of Colorado Boulder, used with permission.

[4] Geographic Information Systems and Cartography by Adam Dastrup under a CC BY-NC-SA 4.0 license

[5] Introduction to Cartography by Ulrike Ingram under a CC BY 4.0 license

Chapter 3
Cartographic Design Process

3.0 The Cartographic Design Process

The cartographic process is the process used to go from map conceptualization to a finalized map. The cartographic process is not a rigid process by which step-by-step directions are followed, however, it provides a recommended set of steps that should be followed in order to properly create a map [7]. Defined, cartographic design is the process of using the principles of design to create a map that has aesthetic appeal while meeting the needs of the map's audience and overall map purpose. As a discipline, cartography integrates design, geography, and technologies such as GIS and remote sensing.

The cartographic process is generally composed of five steps. Step one is to define the purpose and meaning of the map, along with its intended audience. Step two is to choose the scale of the map. The third step is to determine the map format, data needs, printing limitations, and the economics of production of the map. Step four is to abstract and generalize the data. Step five is to finalize the design and layout of the map [7]. However, user feedback is an important part of the design process and can be incorporated into any step of the design process, but at a minimum feedback should be obtained before the map is finalized for production.

Once you reach step five you will often need to make changes, evaluate those changes, and possibly go back to one of the previous steps. Also, it is important to note that the cartographic process does not need to be followed in this particular order. Designing a map is often a dynamic and iterative process which requires you to move back and forth between steps in order to produce a cartographically sound map. We will now go over each one of the five steps in the cartographic process in more detail [7].

Step 1 - Define the Map Purpose and Audience: Defining the purpose and meaning helps the map mapmaker determine the requirements for the map that will meet the needs of the map user. Two questions can be asked to help with these determinations [7]:

1) "what needs to be communicated?": A thorough understanding of what is to be conveyed on the map, who the map user is going to be, and what the user can gain from viewing the map.
2) "how to best facilitate the communication of ideas?" Once the map maker has determined what needs to be communicated they then have to determine the best way to communicate that information. That requires considerations for what format that map will be (tangible or virtual map), what category of map will be created (e.g., thematic), and how the map will be produced and delivered. [7]

Step 2 - Choose Scale: Once the map purpose and user needs have been determined, the map maker must define an appropriate map scale (map scale being the ratio of map distance to distance on the ground). In other words, it is the ratio at which the feature being mapped has been reduced to fit on the map. Map scale operates along a continuum from large-scale to small-scale. For more information on map scale, see Chapter 7. The selection of scale is possibly the most important decision a cartographer makes because scale dictates the look of your map, what scale data will be required, and how much detail can be shown on the map [7].

Step 3 - Data, Format, Printing, and Economics: Quite a few questions must be answered in this step. What kind of map will be created such as a reference or thematic map? What data is needed to create this map and where/how can the data be obtained? How will the map be displayed and accessed? How much time will it take and will it cost to create the map? Depending on your format restrictions, whether you are going to use color or not, etc., all factor into the cost of creating a map [7].

Step 4 - Abstract and Generalize: Once the map purpose, scale, format, and type have been determined, the map maker now has to appropriately abstract and generalize the data to match those criteria. This includes critically analyzing every piece of data used on the map to determine whether it is useful to the map reader and the purpose of the map [7].

As abstractions are made to match the map requirements the cartographer is looking to appropriately reduce the amount of detail in the data so that the map will not contain unnecessary details in order to avoid the map appearing busy, cluttered, or difficult to use. Broadly speaking, there are four methods to abstract and generalize data for a map which are described below [7].

Simplification: Simplification is reducing the amount or detail of the information/data through selection, classification, and/or symbolization. This may include generalizing (e.g., simplification) the geometry of a dataset (Figure 3.0). This can be done manually or through the use of algorithms. See Chapter 8 for more information on generalization techniques [7].

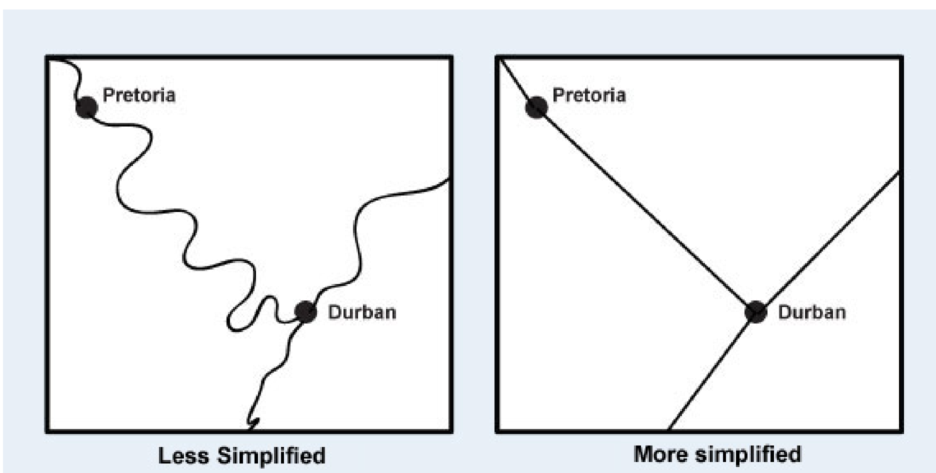


Figure 3.0. Road feature simplification.

Credit: Simplification, *Geographic Information Technology Training Alliance*, Stern et al., CC BY-NC-SA 2.5

Selection: Selection is the process of subsetting a dataset to match the map purpose. For example, if you have a dataset of all of the roads in Colorado, but you are creating a general reference map of the state, you likely only want to show the major roads and highways [7].

Classification: Classification is when observations are put into groups or classes to simplify the data for display on the map (Figure 3.1). In classification, similar map objects or those with similar data values are grouped together, thereby reducing complexity and increasing the ease of communicating the data.

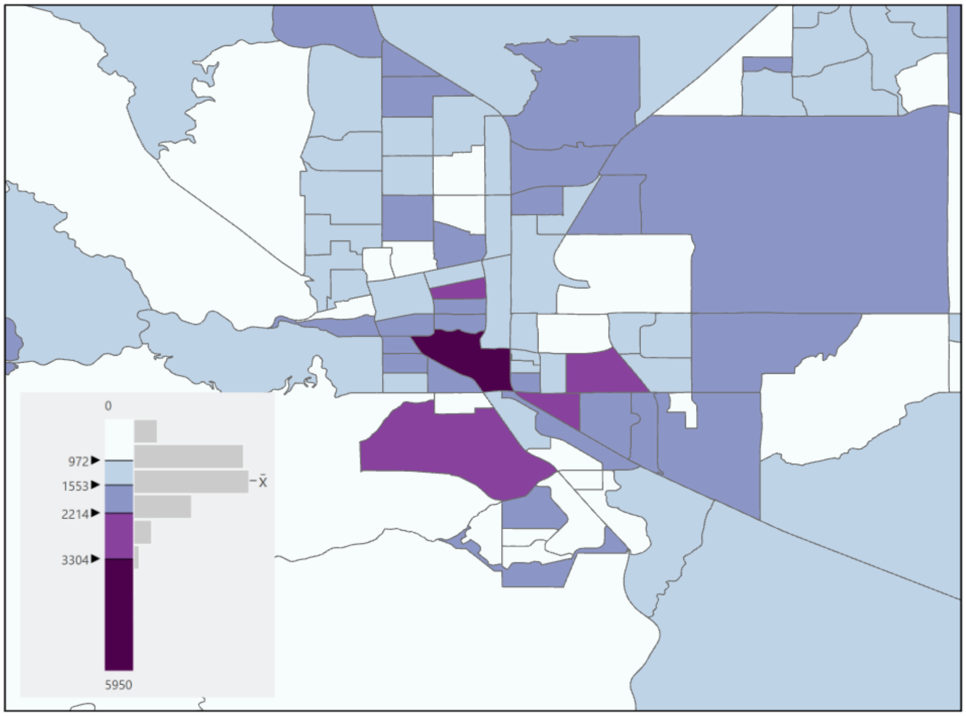


Figure 3.1. Map with classified data.
Credit: Sarah Schlosser, University of Colorado Boulder

Symbolization: Symbolization is when representative shapes or objects are used to represent items or spatial phenomena on a map. Figure 3.2 shows examples of symbols used for the three types of geometric data used in GIS, points, lines, and polygons.

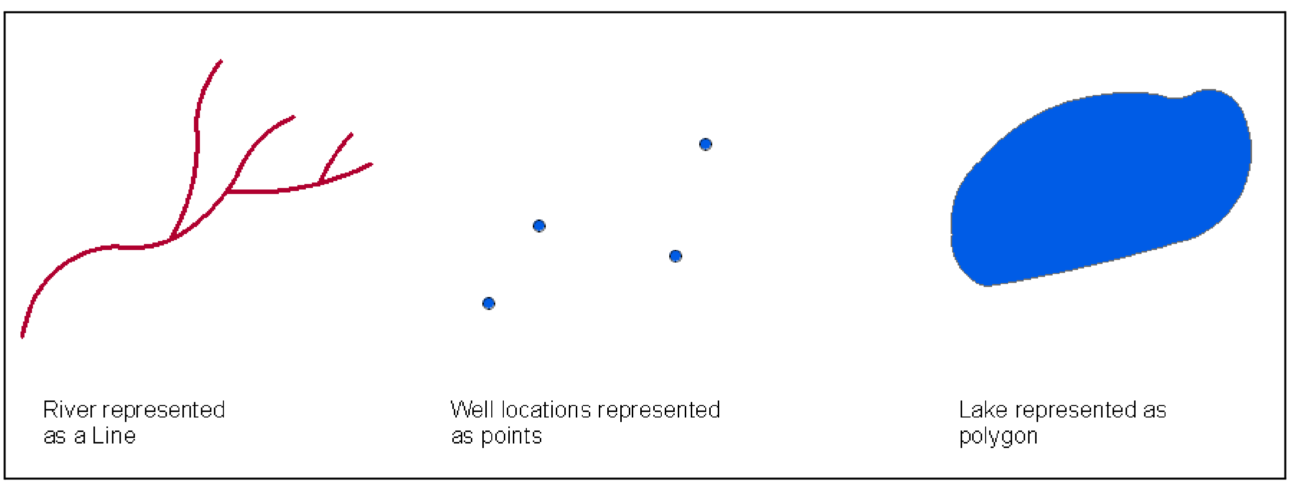


Figure 3.2. Cartography symbol examples.
Credit: GIS PointsLinesPolygons, *Wikipedia*, public domain

When choosing symbols for your map, there are a few things that you should keep in mind. Use universally accepted and understood symbols when possible. For instance, if you want to show the location of an airport on a map, a universal symbol is an airplane. If instead of an airplane you chose to represent it with an airport terminal, this may not be instantly recognizable to the user and will require them to visit the legend unnecessarily. Also, use universally accepted colors when possible. For example, parks are typically green, therefore the map will be easier to read and the feature more readily identifiable to the reader if standard color conventions are used [7].

Step 5: Design the Map Layout: When designing your map layout do not stick to a rigid idea of what the map should look like - always experiment with the map design. Furthermore, rarely should the first draft of a design be used. As indicated above, the cartographic design process is iterative and will go through many iterations before the final design is reached. Feedback should be sought throughout the design process, not only during the layout design phase. Never be afraid to start over from scratch if a map design is not meeting the map's purpose or user needs. During the design process, the map maker should consider whether the map is clear, legible, and aesthetically pleasing (more on these in Section 3.2), seeking feedback on these areas specifically [7].

3.1 Alternative Approaches to the Cartographic Design Process

The cartographic design process outlined in Section 3.0 is only one of many ways in which map design can be approached. Below is another cartographic design process diagram (Figure 3.3). The cartographic process shown below begins with a real or imagined environment. As map makers collect data from the environment (through technology and/or remote sensing), they use their perception to detect patterns and subsequently prepare the data for map creation (i.e., they think about the data and its patterns as well as how to best visualize them on a map). Next, the map maker uses the data and attempts to signify it visually on a map (encoding), applying generalization, symbolization, and production methods that will (hopefully) lead to a depiction that can be interpreted by the map user in the way the map maker intended (its purpose). Next, the map user reads, analyzes, and interprets the map by decoding the symbols and recognizing patterns. Finally, users make decisions and take action based upon what they find in the map. Through their provision of a viewpoint on the world, maps influence our spatial behavior and spatial preferences and shape how we view the environment. In the cartographic design process as diagramed below, the fundamental component in generating a map to depict the environment is itself a process – the process of map abstraction [3].

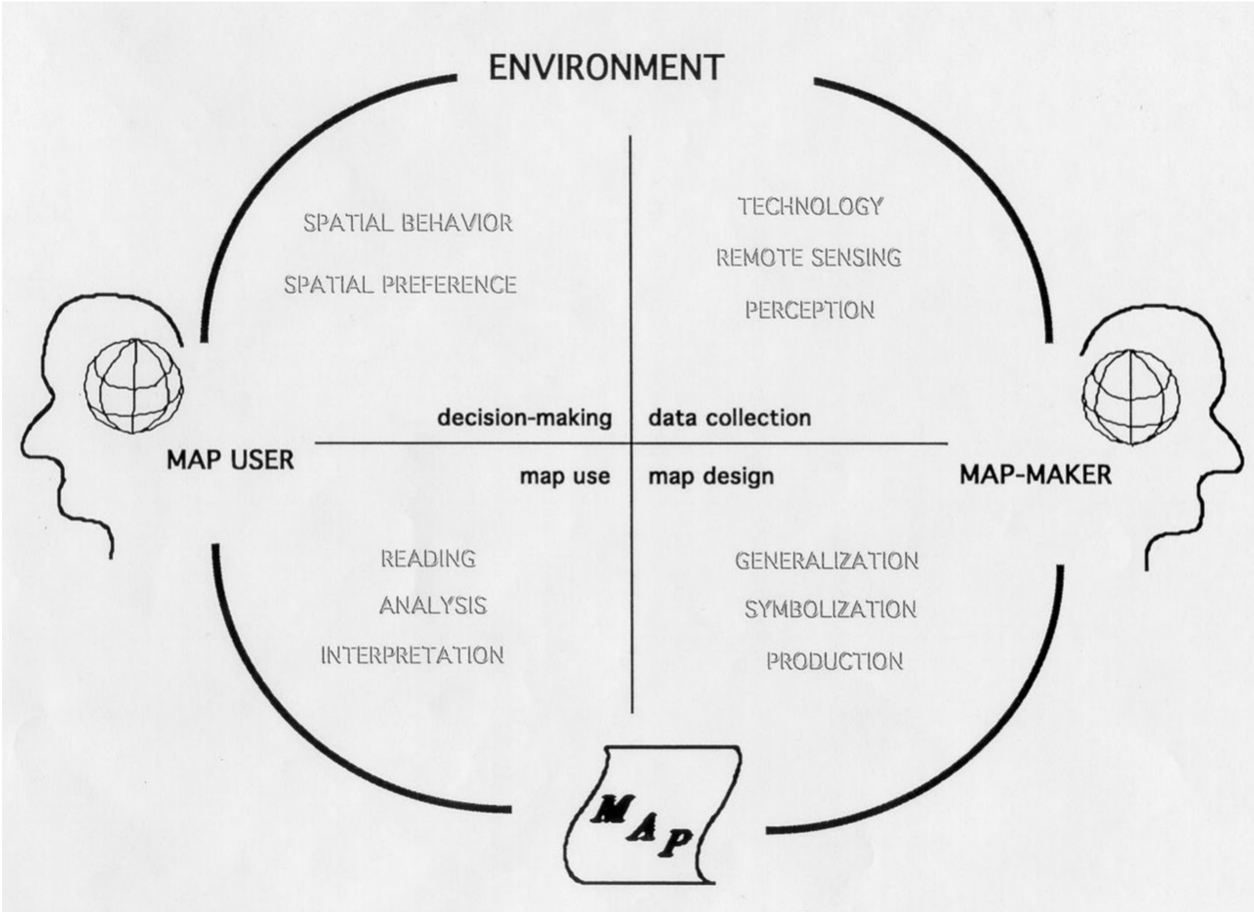


Figure 3.3. Cartographic Design Process
Credit: Cartographic Design Process, Barbara Buttenfield, University of Colorado Boulder, used with permission

3.2 Design Considerations

The goal is to design a map that is a functional visual communication device, that is visually pleasing and accurate. This section focuses on the ways in which the cartographer can achieve that.

Harmony, Legibility, and Clarity: Broadly speaking a cartographically well-designed map will be harmonious in its presentation, legible, and clear. A map is harmonious when the map elements look like they belong together and complement the map's purpose, which means all map elements have a similar feel and style. No one element looks to be out of place. A well-designed map also is legible. This speaks to the map's readability, including the logical association between a symbol and its label. Ideally, the positioning of the map elements will be in accordance with their importance on the map and will be easy to reference relative to the main map body. Lastly, a well-designed map has clarity. This means that the map reader can instantly recognize elements on the map and can easily recognize and use all the elements on the map [7]. This also includes matching the visual priorities (visual hierarchy) with the map message [4].

To accomplish harmony, legibility and clarity, the mapmaker uses the following:

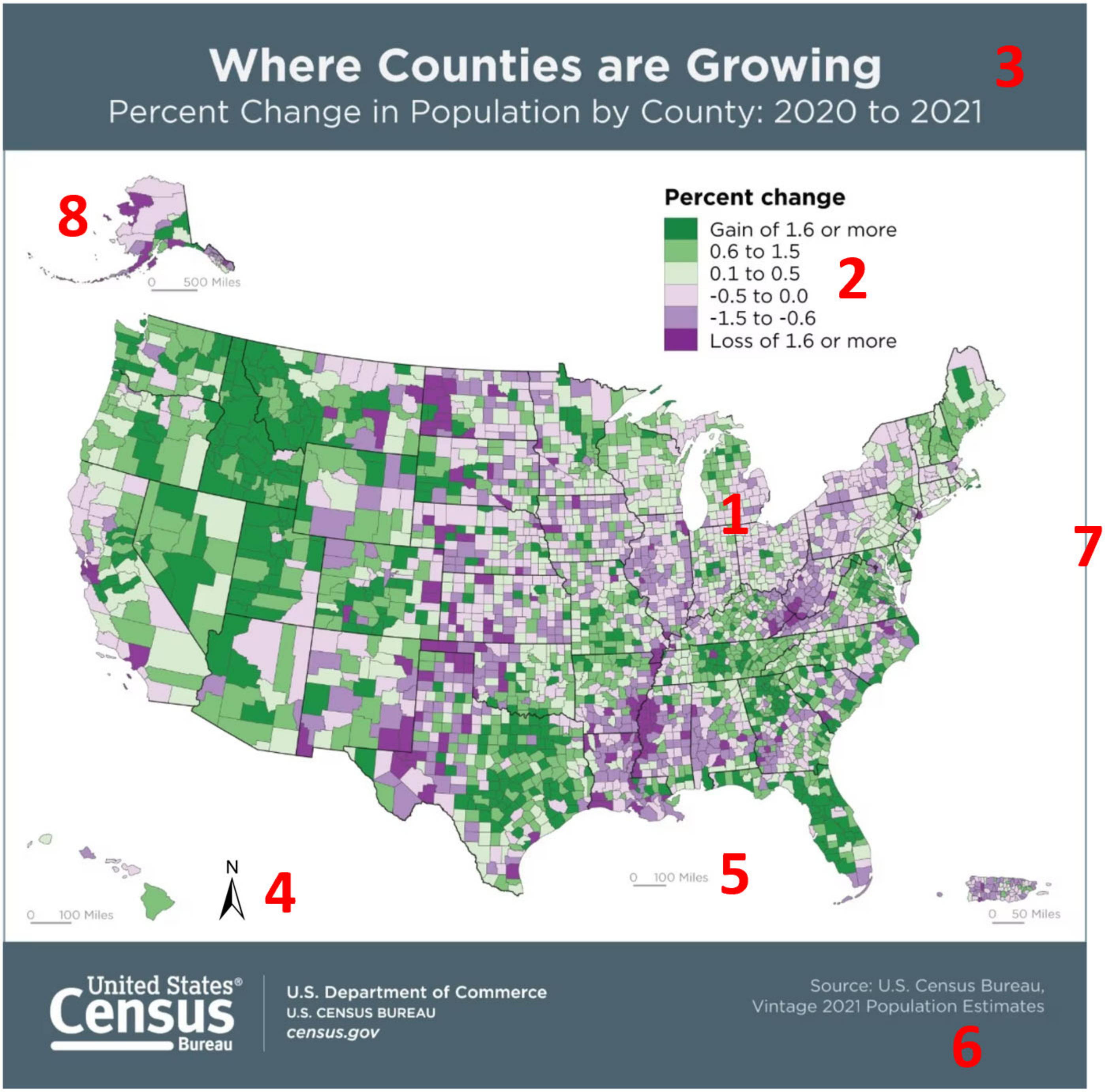
- Generalization: clarifying complex shapes at smaller scales
- Symbolization: creating a logical association between the symbol and its meaning
- Color: colors used are meaningful, pleasing, and balanced
- Typography: labeling is legible and is clearly linked to a named feature
- Layout: composition of map elements is visually balanced

Wright, in 1942 stated this about map aesthetics: “The quality of a map is also in part an aesthetic matter. Maps should have harmony within themselves. An ugly map, with crude colors, careless line work, and disagreeable, poorly arranged lettering may be intrinsically as accurate as a beautiful map, but it is less likely to inspire confidence.” What Wright is alluding to is that for a map to be effective in the eyes of a map user, the map must not only be accurate, but it must also have harmony, legibility, and clarity [7].

3.3 Map Design and Map Elements

A map is composed of multiple parts known as map elements. Map elements that may be on a map include a neat line, map body, graticule, insets, title, legend, label, ancillary text, scale bar, directional indicator, and metadata. It is important to note that not all map elements are required for every map [7].

In Figure 3.4, the map elements are indicated by numbers one through eight with each item listed below the map, ordered from most important to least important to have on a map layout. Each element is discussed in more detail below [7].



1. data (or map) frame
2. map legend
3. map title
4. north arrow
5. map scale bar
6. metadata (or map citation)
7. border
8. inset (or locator) map.

Figure 3.4. Map showing the elements that can be part of a map layout.
Credit: adapted from *Where Counties are Growing*, US Census Bureau, public domain

Data / Map Frame: The map body is the main focus of the map and should be visually dominant. The map body contains the features important to the message of the map and should be as large as possible on the medium, leaving room for other map elements and some "breathing room" around the edges of the map. The map body is the only map element that is absolutely required when composing a map. However, the map is often much more useful when other map elements are included [7].

Legend: The purpose of the legend is to identify map features succinctly. It does not need to include a representation of map features that are self-evident such as water features or roads. However, general reference maps traditionally define all symbols. Symbols that represent map features on the legend should be identical to the symbols on the map [7].

When designing a legend you should not include the word “legend” as a map reader knows that it is a legend. Instead, include meaningful text at the top of the legend or no text at all if it is not needed. An exception to this is if the audience is not skilled at map reading such as young children [7].

Title: The goal of the title is to bring attention to the purpose of the map. It should be dominant in size, brief yet focused, and will typically include *where, what, and when* as related to the map’s purpose. The map title should be easy for the map user to read quickly and with ease. While every standalone map should have a map title, a common reason not to put a title on the map is if it is a figure in a larger body of text and will have a caption. In this case, place a map title in the caption text. A bounding box around the title can be used if the legibility of the title is decreased due to elements on the map. The following are examples of map titles [7].

North Arrow / Compass Rose: The directional indicator is typically a north arrow (Figure 3.5) or compass rose. The purpose of a directional indicator is to indicate the direction of the map. Map readers are used to most maps showing north as up, which if the map is oriented this way, may not need a north arrow, but depends on the map type and purpose. North arrows are necessary on all maps when the north is not at the top of the map/page/screen and where map readers are not familiar with the area shown on the map [7].

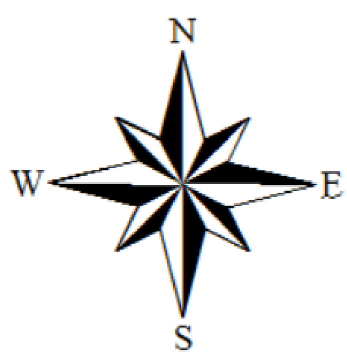


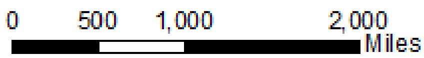
Figure 3.5. North arrow example.
Credit: Directional Arrow, *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

The north arrow should not be made overly large and should not be placed in prominent locations; it should be low in the visual hierarchy. If a map reader needs to know the orientation of the map, they are expected to search for the north arrow. However, if the map is not oriented with the north at the top, the north arrow should be higher in the visual hierarchy to bring attention. When choosing a north arrow make sure it is not overly ornate and easy to use and understand [7].

Scale: The purpose of including scale information on a map is to understand the relationships between features on the map and the features on the ground and to measure distances on the map. Map scale should use a unit of measure that is appropriate for the audience and its purpose (e.g., 1 mile instead of 5280 feet). Additionally, the map scale should be subtle and small, not more than a third of the size of the mapped feature, but not so small as to be of no use [7].

A map scale is typically included on a reference map, but seldom on a thematic map since readers are not expected to measure distances on such maps. When choosing and designing a map scale, use round, meaningful numbers on your map scale; decimals should be avoided on a map scale unless the decimal number is a meaningful and significant interval that you expect map readers to measure (e.g., 0.5 miles). There are three types of scales that you can use for a map: graphic, verbal, and representative fraction (Figure 3.6). These are discussed in more detail in Chapter 7.

Graphic Scale / Scale Bar:



Verbal Scale: One inch on the map equals twenty feet on the ground

Representative Fraction: 1:24,000

Figure 3.6. The three types of scale used on maps.
Credit: Graphic Scale, *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0 (graphic scale only)

Metadata: Metadata is data about the map and its source data. It can include the map author's information, such as their name and contact information. Metadata should also include the date the map was created and other relevant explanatory information about the creation of the map such as the map projection used. It is typically placed along the bottom edge of the map and deemphasized as it is typically not that important for the map reader to notice right away. The metadata should be among the smallest text on the map [7].

Border / Neatline: The difference between a map's border and neat line can be confusing. A border line encloses all elements on the map layout, while a neatline crops the map area. A border might also act as a neat line if all elements (e.g., legend, title) are shown within the map area [5].

Inset Map: An inset map element is a small ancillary map that is typically larger in scale than the map body. The purpose of the inset map is to show an area in more detail, to clarify areas where features would otherwise be overcrowded if restricted to the primary mapping area, or for areas that are geographically farther away (Figure 3.7) such as Alaska and Hawaii that are placed as inset maps. In this case, typically, Alaska is shown at a reduced scale and Hawaii is shown at an enlarged scale. It is also common to see inset maps of the New England area when looking at a map of the United States because there is too much detail to be shown at a map scale that shows the entire United States [6][7].

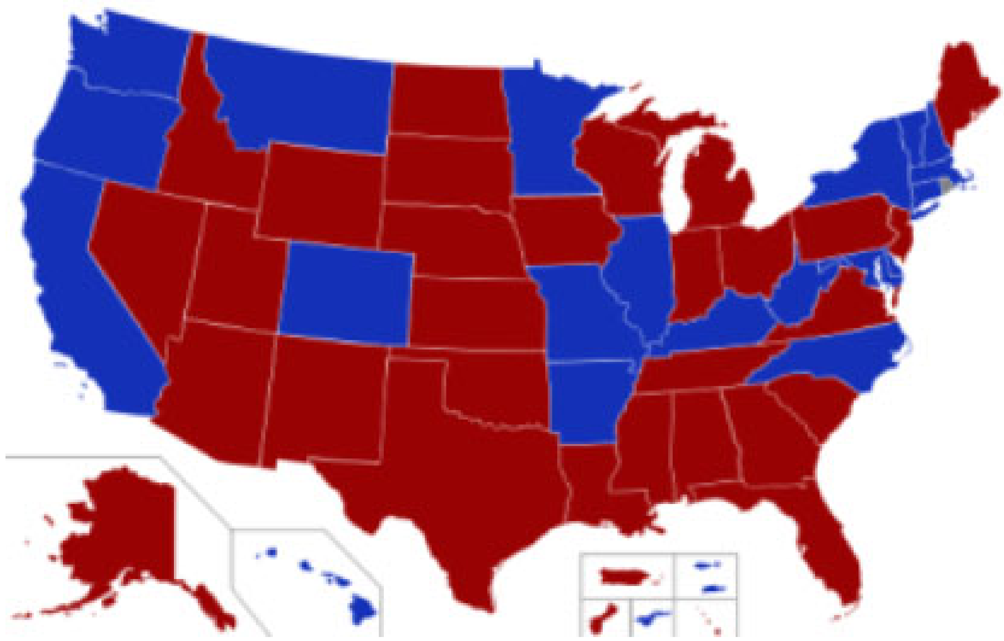


Figure 3.7. The United States is shown with multiple inset maps.
Credit: Inset Map, *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

Locator Map: A locator map is a small ancillary map that is on a smaller scale than the map body (Figure 3.8). The purpose of the location map is to identify the location of the main map body in relation to its larger geographic context. A locator map should be included in a map layout for an area that is unfamiliar or non-intuitive to the map reader(s) [7].

Figure 3.8. Map of Arkansas with a locator map.
Credit: Location map, *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

Graticule: A graticule is a visual representation of a coordinate system used on a map (Figure 3.9). You should include a graticule if the map reader will be referencing coordinate locations when using the map. The graticule should use meaningful divisions and units. Typically the graticule is omitted from a thematic map as thematic maps wish to convey the spatial distribution of a theme and are not to be used for determining precise locations. Graticules should be lower in the visual hierarchy on the map, yet be easy for the map reader to use across the entire map [7].

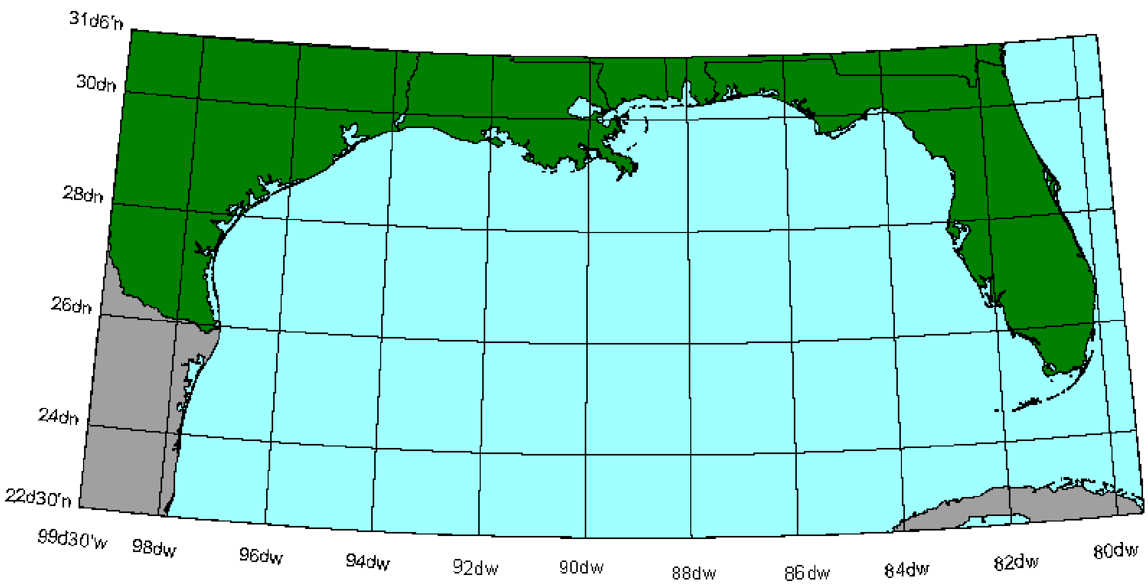


Figure 3.9. Graticule shown on a map.
Credit: Graticule, USGS, public domain

3.4 Legend Design

When designing a legend care should be taken to make it look well-structured and consistent. There are a few particular issues that warrant extra attention while designing your legend. The symbols should be to the left with the description/text to the right. The symbols should be evenly distributed and spaced from each other and the description; spacing and alignment consistent all around (Figure 3.10). The symbols should be aligned with the descriptions to its right with the symbols horizontally centered. Symbols should be of the same size or length based on the type of feature (e.g., polygon, line, point). Last, the descriptions should be left justified, thereby in line with each other. If a map reader notices any irregularity in the placement or alignment of the legend items, it may diminish the map reader's trust in the map [7].

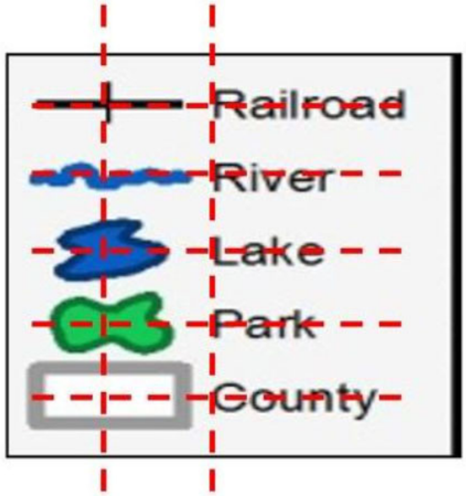


Figure 3.10. Legend Design and Item Alignment
Credit: Legend Design, *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

Legend Columns: When designing your legend you can choose to split the legend into multiple columns so that the map legend is to be placed properly on the map or the legend entries can be split into meaningful separate columns (Figure 3.11). For any decimal numbers on your legend that are smaller than one, there needs to be a leading zero. Ranges of numbers are normally separated by a hyphen or the word "to", however, the word "to" can help reduce confusion when there are negative numbers in the legend [7].

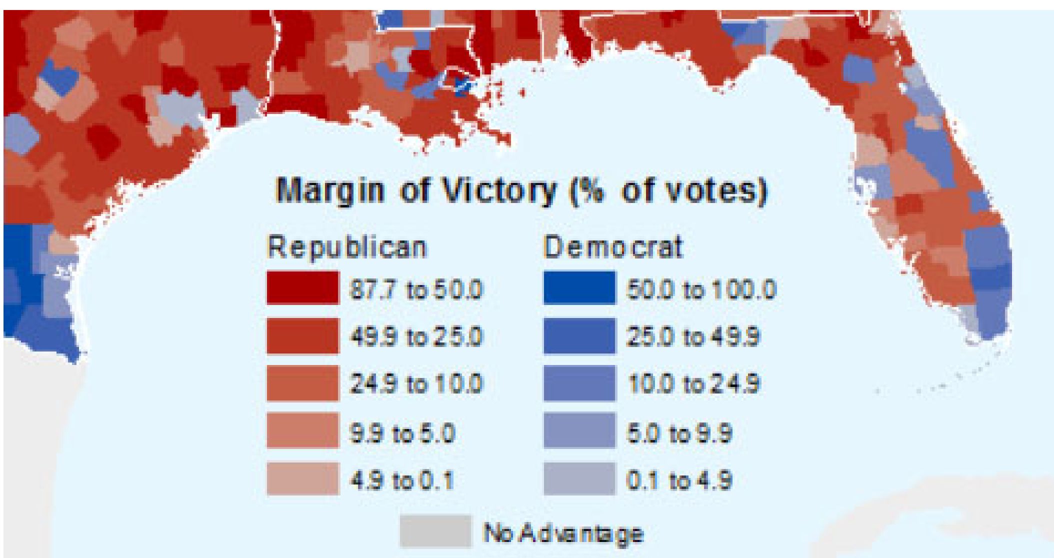


Figure 3.11. Multiple column map legend example
Credit: Columns, *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

Headings: A legend may contain headings to meaningfully separate groups of legend entries. Symbols representing a single feature on the map should have a singular description, while symbols representing multiple features should have a plural descriptionn [7].

3.5 Visual Hierarchy

Not all information on a map is of equal importance. Visual hierarchy is the design principle that takes these factors into consideration. A map with a good hierarchy emphasizes important information and figures by positioning them strategically on the map and by using visual variables (see Chapter 4) appropriately [8].

Map Layout Hierarchy: When planning the hierarchy for the map layout, there are some customary positions to consider. Assuming that a map has five visual levels, visual level one being at the top of the visual hierarchy and immediately viewable and visual level five being deemphasized, something map users may have to search for briefly (Figure 3.12) [7].

On visual level one, the map reader should immediately see the thematic symbols that create the map. Next, the map reader should easily see the title, legend, other symbols, and feature labels on the map. On level two of a visual hierarchy, the user should become aware of the base map which may include land areas and water features. In level three of the visual hierarchy, the map user may notice the scale, graticule, inset map, and north arrow. While these items are important to the overall map layout, they are considered aids for reading the map so they do not need to be readily apparent. On visual level four is the metadata, which is deemphasized, the smallest text on the map, and which is typically placed along the bottom edge. Finally, on level five is the neat line. With the neat line, it is extremely important to help focus the reader on the center of the map. The neat line itself should not draw any attention [7].

Visual Level	Object
1	Thematic Symbols
1	Title, legend, symbols, and labels
2	Base map - land areas: political and physical
2	Base map - water features
3	Scale, graticule, inset map, north arrow
4	Metadata
5	Neatline

Figure 3.12. Map layout visual levels.
Credit: Visual Levels, *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

Visual Contrast: Contrast can assist in defining the internal organization on our map and leads to the differentiation of objects in our map. It can help the map reader distinguish between important and unimportant parts of a map as well. Contrast works differently in color and in monochromatic design, so the map design will need to be adjusted accordingly. Contrast can be achieved through a variety of methods including size, value, and position [7]. See *GIS&T Body of Knowledge* https://gistbok.ucgis.org/bok-topics/visual-hierarchy-and-layout [opens in new tab] for a comprehensive list and visualizations of the ways in which contrast can be applied to a map.

Figure and Ground: Figure-ground takes contrast one step further by looking to build visual "layers" of map information, with items lying above or below other items. It is based on Gestalt Psychology, where the eye and mind separate the Figure (forward) from the Ground (back) in any display automatically. The Figure is perceived as a coherent form with clear outlines (contour) in front of surrounding background information. In designing a display, the map maker must decide what is Figure and make it prominent. On each map or map layout there can be multiple Figures and levels of Figure possible [4]. Grounds are considered to be unbroken behind the Figure so there is no need to make the Figure semitransparent unless the ground is significantly obscured by the figures [7].

To achieve good Figure and Ground, the map maker can add differentiation and/or detail between features. Other key techniques are to make the Figure a closed form, use familiar shapes for features, increase contrast (e.g., lightness on a dark background), put small items "on top", center items in the display, and maintain good "contour" by using clear outlines [4].

Figure and Ground example: In Figure 3.13 there is a strong figure-ground relationship. The United States has a strong outline that allows the eye to focus on the country. The graticule is considered to be ground, has weaker lines, and is interrupted by the United States polygon, but the map reader can assume that the graticule continues uninterrupted by the figure [7].

Figure 3.13: Figure and Ground Relationship

Credit: Figure and Ground Relationship, *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

Visual Balance and Stability: Defined, balance is the visual impact of the arrangement of the map elements. There are two factors to consider related to the balance: weight and direction. Weight refers to the location, size, and shape of the map element of the map composition. Direction refers to the relative location, shape, and subject of the map elements on the map composition [7].

On any image, and specifically a map, there are two centers of the image (Figure 3.14). The first center is the geometric center. The geometric center is the location that is exactly half the height and half the width of the image. The optical center is a location that the human eye initially rests on when viewing an image. The optical center is slightly above the geometric center of the image. This is important to know for map composition as it is the first thing the map reader views and helps the map maker create a balanced layout. Therefore, it is going to be important that you only place the most important features of your map in the optical center [7].

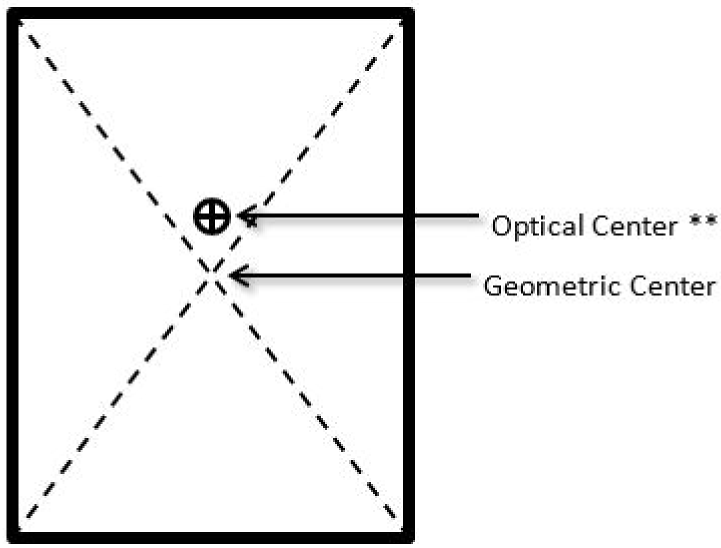


Figure 3.14. Optical (visual) and Geometric Center.
Credit: Map Centers, *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

When considering weight as related to balance on a map, the visual weight of an object depends on its location in the map layout. Elements at the center of the layout pull less weight than those off-center and off-track. An object in the upper part has more weight than in the lower part of the layout. Objects on the right are considered to be heavier than the objects on the left and weight increases in proportion to the distance from the center. If the center of the map does not have as much weight as elements placed around the edges, or one edge of the map has significantly more elements on the other edges, the map would feel off-balance to the map reader. Visual weight also depends on the size of the object. Large objects are heavier than small objects. If a large object is placed off-center and off-track it will have significant weight and may run the risk of throwing the entire map off-balance. Because of all of this, it is important that we evenly distribute the weight around the map [7].

Visual weight also depends on color. Related to color red is heavier than blue, white is heavier than black, and bright colors are heavier than dark colors. And finally, visual weight depends on the shape. Regular-shaped objects are heavier than irregular-shaped objects. Compact shapes are heavier as well. Therefore solid geometric objects such as circles, squares, and triangles are heavier than other geometric shapes such as stars [7].

The visual direction depends on location (Figure 3.15). The weight of an element attracts neighboring objects imparting a directionality to them. The visual direction also depends on the shape. Shapes of objects create axes that impart directional forces into opposing directions. Objects possessing intrinsic directional forces impart visual direction on other elements [7].

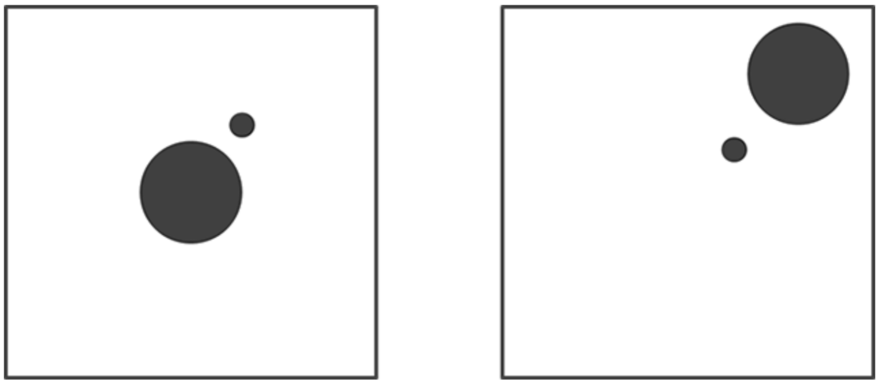


Figure 3.15. No Direction (Left); Upper Right Direction (Right)
Credit: Image from *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

Balancing White Space: Related to balance is the use of white space on the map. White space refers to blank spaces on the map that do not contain an element. In general, white space should be minimized in a map, however, white space can be useful to visually separate or group elements. Excessive amounts of white space can also make the map visually unbalanced. Consider the white space around South America in these nine images in Figure 3.16. The center image has South America centered on the visual center. Because of this, even though there is as much white space in all of the images, when South America is placed at the visual center, the map layout interrupts the white space and provides balance to the layout [7].

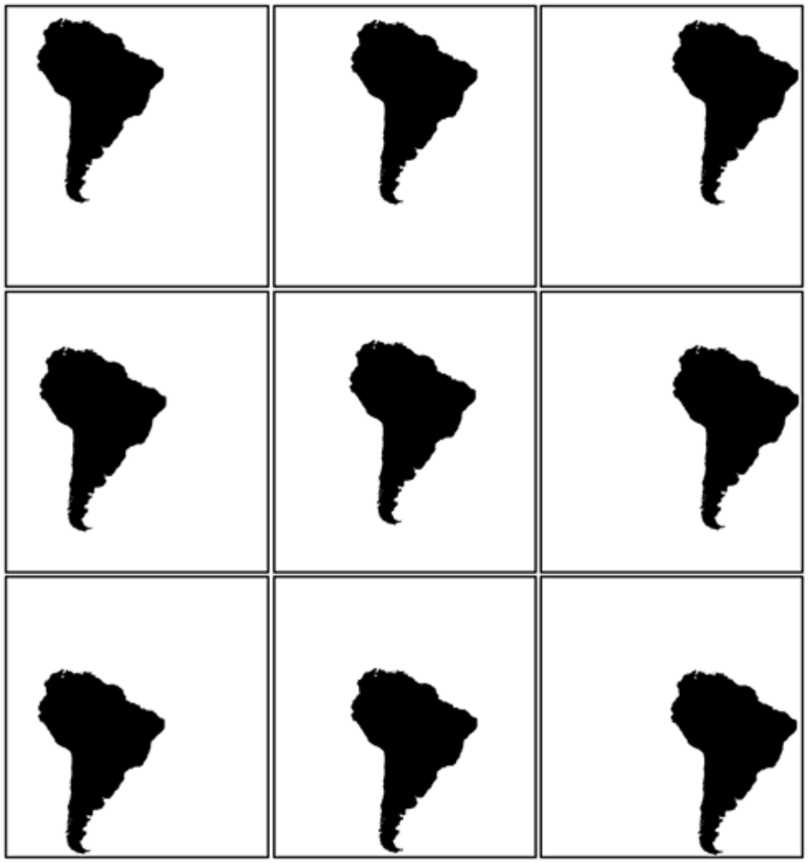


Figure 3.16. White space and balance. The centermost map has the best balance.
Credit: White balance, *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

Figure Closure: Another concept related to visual hierarchy is figure closure. Figure closure is the tendency for the viewer to complete unfinished objects or objects that may be obscured by other elements. For example, in Figure 3.17, the image illustrates how the San Antonio label obscures multiple segments of outlines for multiple counties. The map reader will assume that the county outlines exist beneath the label and will consider those items to be closed figures [7].

Figure 3.17. Showing San Antonio label over the closed figure county boundaries.
Credit: San Antonio Label, *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

Strong Edges: Crisp edges help to define an object as a figure and visually set it apart from other objects. Reducing the edge definition of an object weakens its visual hierarchy. Edges result from a contrast of brightness, texture, or line type [7].

In Figure 3.18, the county boundaries are well-defined as they have strong solid edges. The circles are also well-defined as they have crisp edges. Where the two circles overlap in the center, the larger circle has been given a white edge so that it is visually distinct from the circle and county outlines that lie underneath [7].

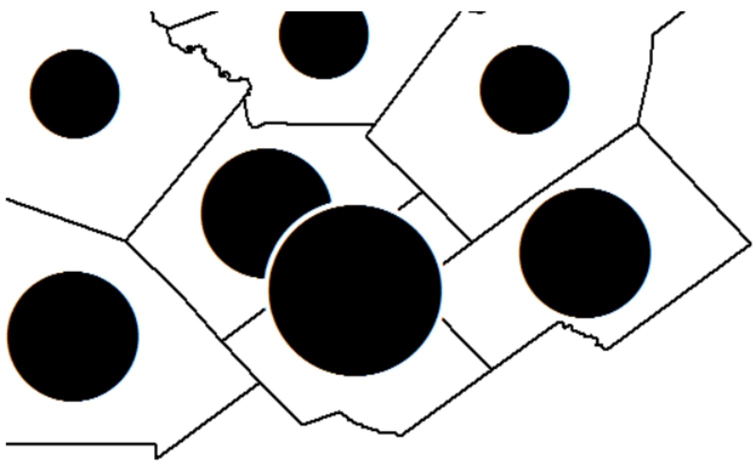


Figure 3.18: County boundaries used to create strong edges.
Credit: Strong edges, *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

Interposition: The final topic related to visual hierarchy is interposition. Interposition is the interrupting of the edge of one object with another object that causes the object to appear on top and closer to the viewer. If two objects overlap, be careful to not overlap them too much. The rule of thumb is that there should be no more than 1/3 overlap between the objects [7].

To illustrate this point consider Figure 3.19, the top square is overlapping the bottom square. As the top square is complete and the bottom square is being interrupted, we visually perceive one square to be “on top” of the other [7].

Figure 3.19: Interposition Example.
Credit: Interposition Example 1: Overlapping Squares, *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

Additional Resources
Visual Hierarchy and Layout: https://gistbok.ucgis.org/bok-topics/visual-hierarchy-and-layout [opens in new tab]

References - materials are adapted from the following sources:

[1] Cartography Guide by Axis Maps under a CC BY-NC-SA 4.0 license

[2] Essentials of Geographic Information Systems by Saylor Academy under a CC BY-NC-SA 3.0 license

[3] GEOG 160 Mapping Out Changing World by Joshua Stevens, Jennifer M. Smith, and Raechel A. Bianchetti, Pennsylvania State University, under a CC BY-NC-SA 4.0 license

[4] GEOG 3053 Cartographic Visualization by Barbara Buttenfield, University of Colorado Boulder, used with permission.

[5] GEOG 486 Cartography and Visualization by Cary Anderson, Pennsylvania State University, under a CC BY-NC-SA 4.0 license.

[6] Geographic Information Systems and Cartography by Adam Dastrup under a CC BY-NC-SA 4.0 license

[7] Introduction to Cartography by Ulrike Ingram under a CC BY 4.0 license

[8] Mapping, Society, and Technology by Steven M. Manson under a CC BY-NC 4.0 license

Chapter 4
Visual Variables

4.0 Levels of Measurement

It is imperative to consider the distinctions between kinds of data, namely whether they are quantitative or qualitative. Qualitative data deal with descriptions of a real-world phenomenon that relate to differences in kind or existence such as types of fruit (e.g., apples, oranges, grapes). Quantitative data are those that deal with measurements (or quantities) that deal with differences in amount such as household income or population density. A qualitative map of cities, for example, would show whether a city exists or not in a given place, while a quantitative map would show the location of the city as well as some measurements, such as the number of people living there [2].

The levels of measurement can be further broken down into nominal, ordinal, interval, or ratio. If the data is nominal, this means that it uniquely identifies items and shows them as different from other items, there is no order or value assigned to the data. If data is categorized as ordinal, there is some type of ranking or hierarchy to the data such as low, medium, or high. Note that the specific data values are not indicated with ordinal data. Interval data is data that occurs along a scale, such as temperature, where a zero value does not indicate an absence of the phenomena being measured. Ratio data is that which occurs on a numerical scale for which there is an absolute zero, such as population counts. Figure 4.0 shows the different types of data, how they differ, and examples of each type. Figure 4.1 shows how each type of data can be portrayed in a map legend [2].

Qualitative		**Quantitative**	
Nominal	**Ordinal**	**Interval**	**Ratio**
Named	Named	Named	Named
	Ordered	Ordered	Ordered
		Proportionate intervals	Proportionate intervals
			Has absolute zero
Examples			
Hair Color, Major, Marital Status	Grade level (freshman, sophmore...); quality rating (low, medium, high)	temperature, time of day	height, weight, distance, salary

Figure 4.0. Types of data.

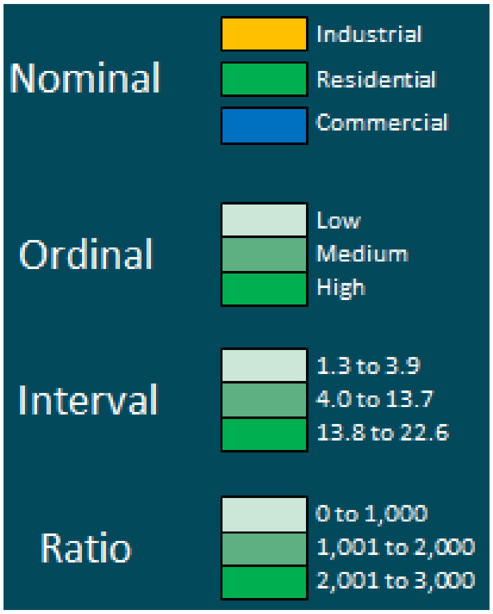


Figure 4.1. Portrayal of data types with example legends.
Credit: Levels of Measurement, Adapted from *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

4.1 Visual Variables

Visual variables are distinctions that are used to create and differentiate symbols on a map. This text focuses on six visual distinctions commonly used for symbolization: size, shape, orientation, texture, color hue, and color value (Figure 4.2). They were developed by Jaques Bertin, and published in the book *Semiology of graphics: diagrams, networks, maps* (1983) [1]. This section will cover each visual variable and will be discussed in the context of whether it is most useful for qualitative or quantitative mapping.

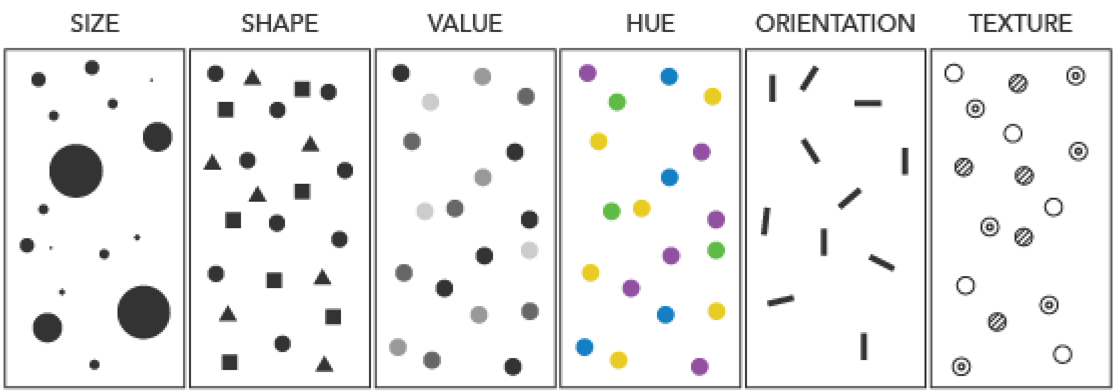


Figure 4.2. Visual Variables.
Credit: Berton's Visual Variables, Adapted from *Cartography Guide*, Axis Maps, CC BY-NC-SA 4.0

4.1.1 Qualitative Visual Variables

Qualitative visual variables are used for nominal data. The goal of qualitative visual variables is to show how entities differ from each other, and to group similar entities.

Hue: Hue (Figure 4.3), more commonly known as color, represents a wavelength on the visible portion of the electromagnetic spectrum. Hue is great for identifying items as unique, or grouping by a type of item (categories).

Figure 4.3. Hue/color.
Credit: hue, adapted from [NASA Earth Observatory](), public domain

Orientation: The orientation visual variable changes the orientation or direction of the object and creates a perception of grouping or likeness (categories).

Shape: The shape visual variable identifies an item as unique or a group of a type of items (categories). The shape visual variable typically refers to a point symbol although it can be arranged to resemble a line and placed inside an area or three-dimensional shape. The shape does not have to be a geometric form (e.g., circle, square), it can also be pictorial (e.g., airplane symbol) [2].

Texture: Texture refers to the areal fill, and its relative courness or fineness, covering an area. Textures identify items as unique or of a type (categories), but can also be used in a sequence of increasing or decreasing coarseness (Figure 4.4). However, using texture with quantitative variables should be done with extreme caution as it can easily confuse the map reader if not correctly done [2].

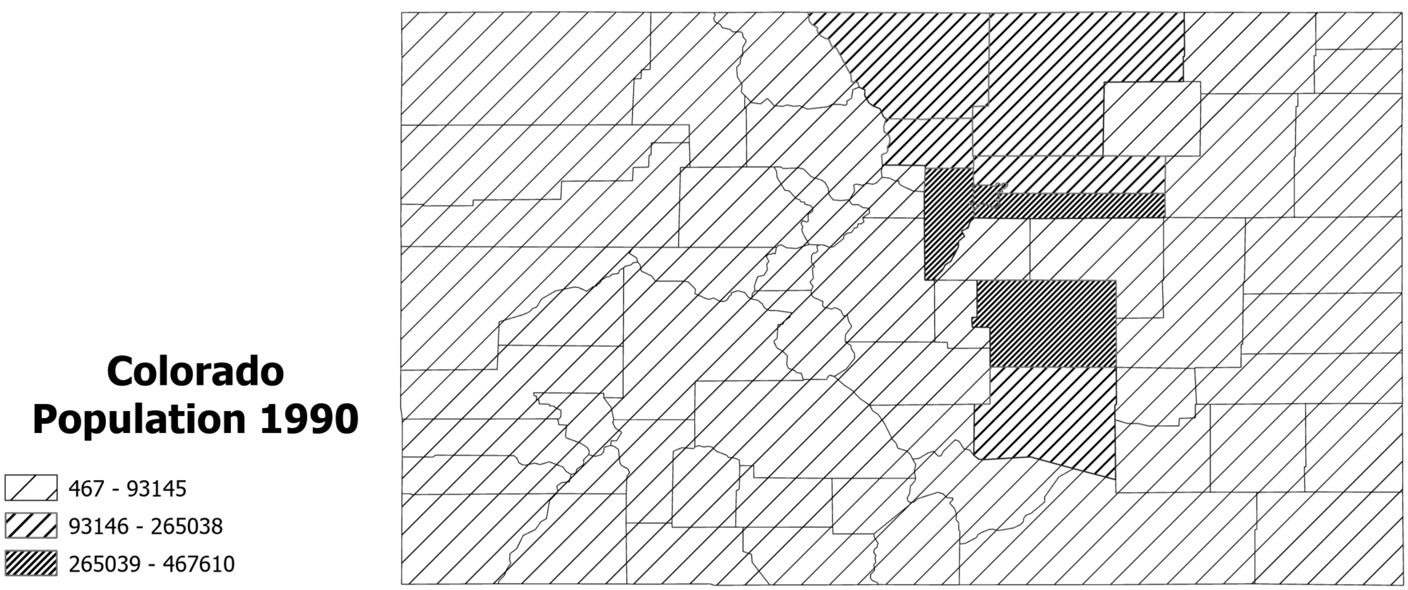


Figure 4.4. Colorado Population total counts; effective application of the texture visual variable for quantitative data by using hatching separation, note all hatching is of the same pattern/direction.
Data Source: US Census Bureau

4.1.2. Quantitative Visual Variables

Quantitative visual variables are used to display ordinal, interval, or ratio data. The goal of the quantitative visual variable is to show the relative magnitude or order between entities.

Size: The size of a symbol can change to imply relative levels of importance or quantity. For example, with graduated symbols a large circle denotes a higher value of the mapped phenomena. For linear features, line thickness implies relative flow levels in the case of road traffic or can indicate water flow through a river (thicker line = more water) [2].

One consideration when using size, in particular with circles, is the Ebbinghaus illusion which occurs when surrounding circles influence the perception of symbol size as is shown in Figure 4.5, where both central circles are the same size, but appear to be different sizes. A way to help reduce this illusion is by using feature borders, such as counties, to interrupt the space between the circles.

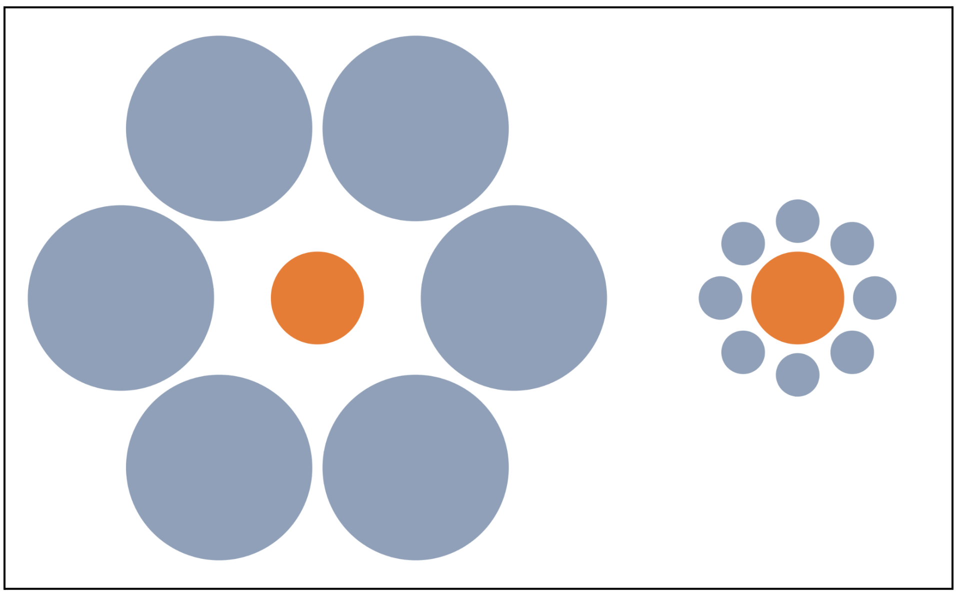


Figure 4.5. Ebbinghaus illusion; both central circles are the same size.
Credit: Mond-vergleich, *Wikipedia*, Fibonacci, public domain

Color Value: The visual variable color value (Figure 4.6), often referred to as lightness, represents different magnitudes or orders of data values. Value is used to represent a single variable by a single hue, with different quantitative values represented by a difference value (lightness of a particular hue). Value can be used in grayscale when hue cannot applied [2].

Figure 4.6. Value/lightness.
Credit: greyscale palette, adapted from NASA Earth Observatory, public domain

4.2 Matrix of Visual Variables

Below is a matrix (Figure 4.7) that matches visual variables with data type, indicating which is best for each type of data. The empty cells represent an incompatible data type and visual variable and should be avoided. The “Good” cells represent a marginal match of the visual variable with the data type and should be used with careful consideration. The “Best” cells represent a matching visual variable with the data type and should be the most common pairing.

Qualitative	Quantitative		Visual Variable
Nominal	**Ordinal**	**Interval and Ratio**	
Good			**Shape**
Good	Good		**Orientation**
BEST			**Color (hue)**
Good			**Texture**
Good	BEST	BEST	**Value (lightness)**
	BEST	BEST	**Size**

Figure 4.7. Matrix matching visual variables with data type.

Additional Resources
Visual Variables - https://gistbok.ucgis.org/bok-topics/symbolization-and-visual-variables [opens in new tab]

References - materials are adapted from the following sources:

[1] Bertin J. (1983). Semiology of graphics: diagrams, networks, maps. University of Wisconsin Press.

[2] Introduction to Cartography by Ulrike Ingram under a CC BY 4.0 license

[3] Mapping, Society, and Technology by Steven M. Manson under a CC BY-NC 4.0 license

Chapter 5
Thematic Symbols

Thematic symbols determine how data on the map is read and interpreted and are a major part of creating a successful map. Using a combination of visual variables (e.g., hue, value, texture) and feature types (point, line, area) allows cartographers to customize and design symbols specific to the data. In this chapter, point symbols will be covered first, followed by lines and areas, and in each section the text covers the ways in which the types of features can be represented [1].

5.0 Point Symbols
There are three types of point symbols: 1) those representing the distribution/concentration of a particular phenomenon, 2) those that depict data occurring at a particular location such as a city, and 3) those that represent values summarized over an area such as a state. Each will be discussed in further detail in this section [2].

5.1 Dot Density Maps
A dot density map is a map showing total values represented by randomly placed dots within an enumeration area to represent density and spatial distribution/concentration [2]. This example map shows the population density for the year 2000 for the state of Colorado. In Figure 5.0, one dot represents 500 people. Each dot is randomly placed within its enumeration unit, which in this case, is the county. Where the dots are denser on the map the user will interpret this area as having more value (higher population density). Where there are fewer dots on the map, the user will interpret this as having less value (lower population density) [2].

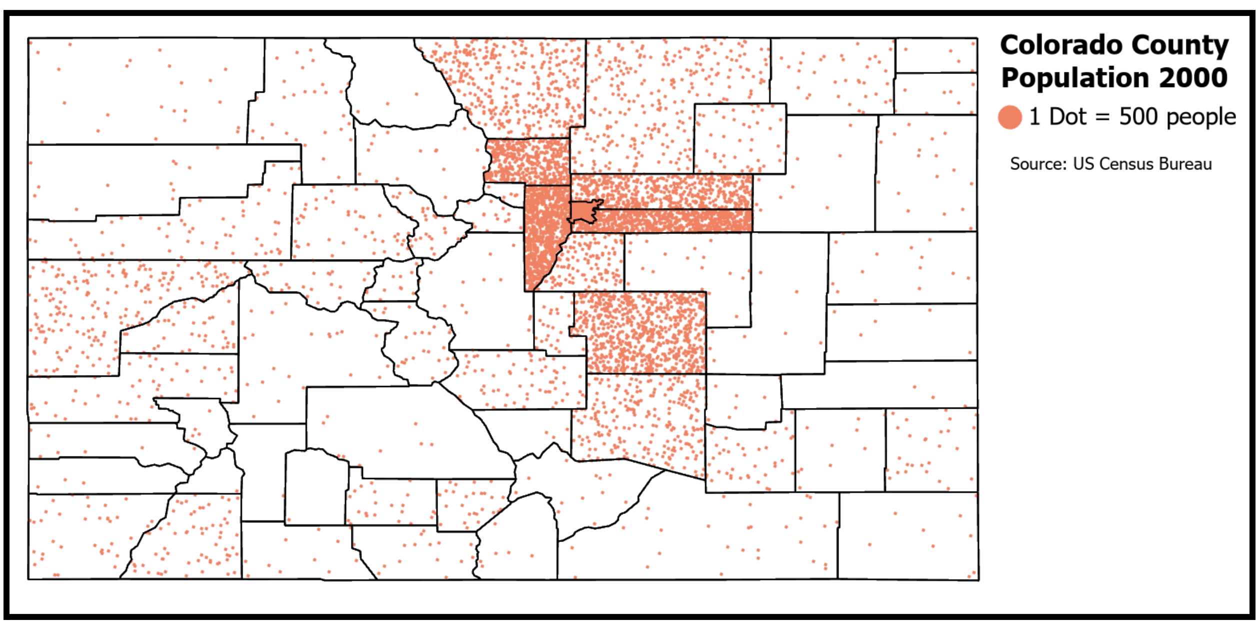


Figure 5.0. Dot Density Map of Colorado Population.
Data Source: US Census Bureau

Positives: A dot density map is easy to create and interpret. It excels at displaying a variable's overall geographic pattern and density. Dot density maps can reflect distributions more accurately than other thematic map types [2].

Negatives: Map readers do not perceive dot densities linearly; they may overestimate or underestimate the densities of the dots as density increases in an area. Second, the dots may not be automatically placed close to the phenomenon they represent since the dots are randomly placed. Third, it may be hard for the map users to recover original totals when dots are placed close together making it hard to pick out individual dots. Finally, dots may appear where they cannot possibly exist. For example, a dot representing cattle may show up in a lake, which does not make sense for the dot to be there. In such situations, dasymetric mapping can be utilized (see Dot Placement below) [2].

Data for Dot Density Maps: Total values that are aggregated to an enumeration unit must be used for a dot density map. The smallest enumeration units possible to maximize the likelihood that the dot will be placed close to the location of the phenomenon is preferred [2]. Generally, the enumeration units themselves are not shown on the map or are done so only to provide geographic context. However, the map should indicate at what enumeration unit the data is represented if the units are not shown on the map. Another option is to show the next level up of an enumeration unit from that at which the data was collected. For example, if the data's enumeration unit is counties, then state boundaries could be shown. Data sets that do not have too large or too small of value ranges are most appropriate for a dot density map. Derived data such as persons per square mile and continuous data are not appropriate for a dot density map [2].

Dot Representation: Dots should be large enough to stand out but small enough that they do not overwhelm the map (Figure 5.1). Enumeration units with the smallest value should have about two to three dots inside of them. Dots should just begin to coalesce in the most dense enumeration area. Dot values should be easy to understand and are typically a round number. For example, 100 and 1,000 are good dot values [2].

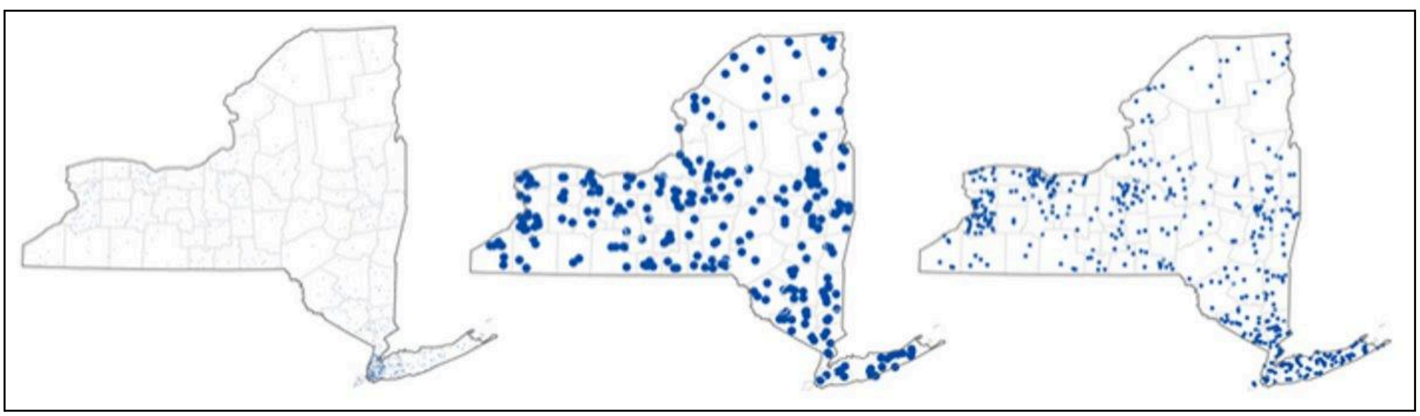


Figure 5.1. Dot size too small (left), dot size too big (middle), and ideal dot size (right).
Credit: dot density maps, Dr. Barbara Buttenfield, University of Colorado Boulder, used with permission

Dot Placement: Each dot is randomly placed inside the enumeration unit to avoid giving the impression that the dots are precisely placed and to avoid regular placement of the dots. Ancillary data can be used to restrict the placement of dots, known as dasymetric mapping [2]. For example, the mapmaker can use land use data to restrict the placement of dots representing people to only areas in which they might live, excluding places such as water bodies and parks. Dasymetric maps are not exclusive to dot density maps (Figure 5.2).

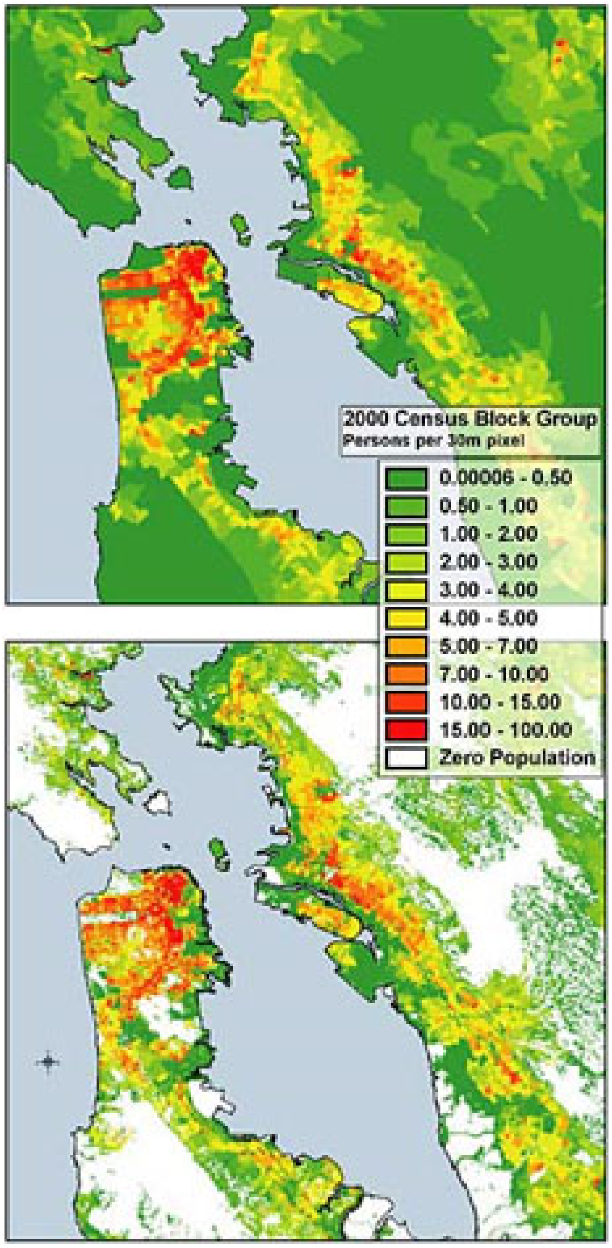


Figure 5.2. Choropeth map (top) and dasymetric map (bottom) of the population in the San Francisco Bay area.

Credit: Dasymetric Map of the San Francisco Bay Area, USGS, public domain

Other Symbols: It is possible to use symbols other than dots. Geometric shapes are familiar to map readers and are simple and easily recognizable. Another option is to use pictorial shapes which may add to the theme of the map. The negative aspect of pictorial shapes is that if the pictorial shapes are too complex, they may not scale well to multiple sizes, thereby the data pattern is lost to the reader [2].

Dot Density Map Legends: On the legend, there should be a representative symbol, a statement about the value of that symbol (1 dot = 100 people), and, optionally, the total value of all of the dots on the map. Additionally, the mapmaker may choose to include representations of what a low, medium, and high density looks like on the map (Figure 5.3) [2]. Dot density maps can be multivariate by using another visual variable, such as hue, to represent a qualitative variable (Figure 5.4) or value/lightness to represent a quantitative variable.

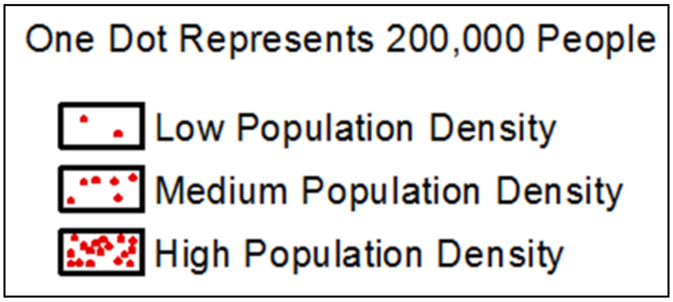


Figure 5.3. Example legend for a dot density map.
Credit: Legend Design, *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

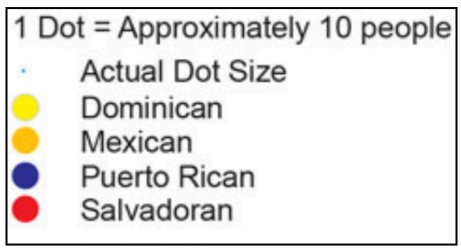


Figure 5.4. Multivariate dot density legend example.
Credit: Static Dot Distribution Map of New York City, adapted from the US Census Bureau, public domain

Map Projections: The best map projection for a dot density map is the equivalent or equal-area map projection. As the relative size is important to maintain when comparing values within enumeration units, a map projection that maintains relative size should be used.

5.2 Proportional Symbol Maps

On a proportional symbol map, a symbol's size is in proportion to the quantity it represents. The most common symbol used in a proportional symbol map is a circle. These maps are effective when the range of data values is too great to utilize a dot map, when magnitudes need to be shown, and when the map needs to represent the quantitative distribution of a variable throughout space. Unlike dot density maps, proportional symbol maps can use normalized data such as ratios or percentages. They can be used to symbolize totals at a point (e.g., city population) or totals over an area (e.g., power generation by region) [2].

Positives: A proportional symbol map is easy for map readers to understand and multiple variables can be displayed simultaneously if desired. For example, the symbol's size, color, and shape can all represent different variables. Additionally, you can overlay proportional symbols onto other types of thematic maps such as a choropleth map [2].

Negatives: Map readers tend to underestimate the relative size of the symbols on the map which leads to errors in estimating values. Also, symbols can easily overlap too much in a dense location, making it hard to see individual symbols and/or determine which entity a symbol is associated with; partially transparent symbols can help alleviate this issue (Figure 5.5) [2].

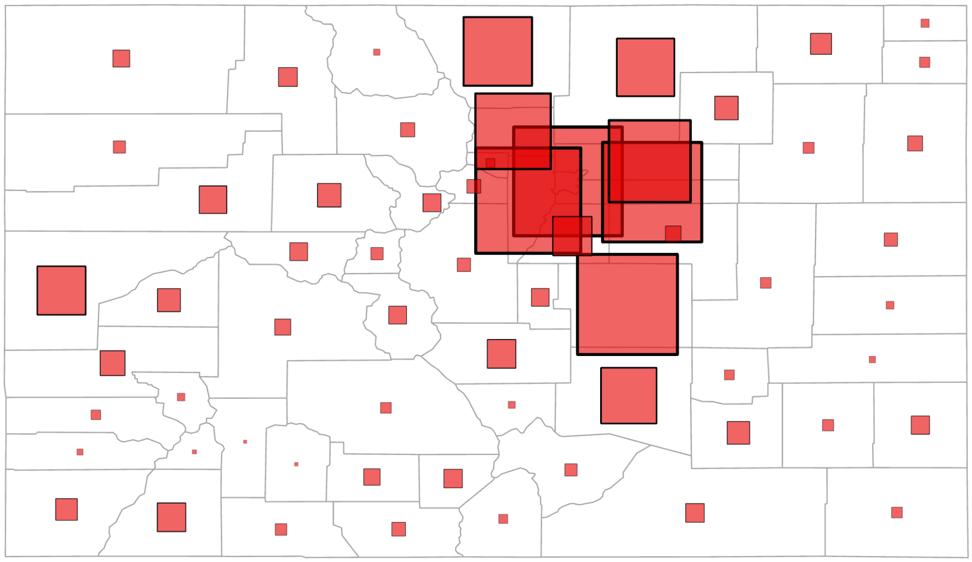


Figure 5.5. Colorado population map using partially transparent proportional symbols.
Data Source: US Census Bureau

Data for Proportional Symbol Maps: Appropriate data for a proportional symbol map is either total values or normalized data such as percentages. Classed data cannot be used for proportional symbol maps - graduated symbols should be used for such data. Also, neither interval data (data that occurs both above and below zero) nor density data are appropriate for proportional symbol maps [2].

The data must occur at a point or be aggregated to an enumeration unit such as a county. If the data is aggregated for an enumeration unit, the symbol is typically placed in the center of the area. Data with small ranges of values will make for an interesting map as the size of the symbols will not vary greatly enough for proportional symbols to effectively portray the spatial phenomena [2].

Proportional Symbols: Various shapes can be used for proportional symbols, although the circle is the most common. There are two categories of point symbols that we can use: geometric symbols and pictographic symbols. A geometric symbol can be either two-dimensional or three-dimensional [2].

Examples of 2-D geometric symbols are a circle, a triangle, and a star. A pictographic symbol looks like the object being mapped, such as a person to represent population counts (Figure 5.6). Pictographic symbols are usually attractive and attention-grabbing, but are harder for the reader to perceive the magnitude differences compared to geometric symbols [2].

Figure 5.6. Colorado population map using pictographic proportional symbols.
Data Source: US Census Bureau

For the 2-D geometric symbol, the area of the symbol is scaled to represent the magnitude difference of the values being mapped. The circle is the most widely used symbol of a proportional symbol map. One advantage of using a circle is that its geometric form is compact and they are visually stable since the eye does not wander across the circle too much. Circles also can easily lend to a second variable by changing its color (Figure 5.7) or texture, or creating pie charts within the circles (Figure 5.8). Squares are also a popular choice (Figure 5.5) with the main advantage being that the proportional areas are nearly perfectly perceived reducing the amount of over or underestimated of the data values. A disadvantage of using squares is that it adds “squareness” to the map which may not be visually desirable [2].

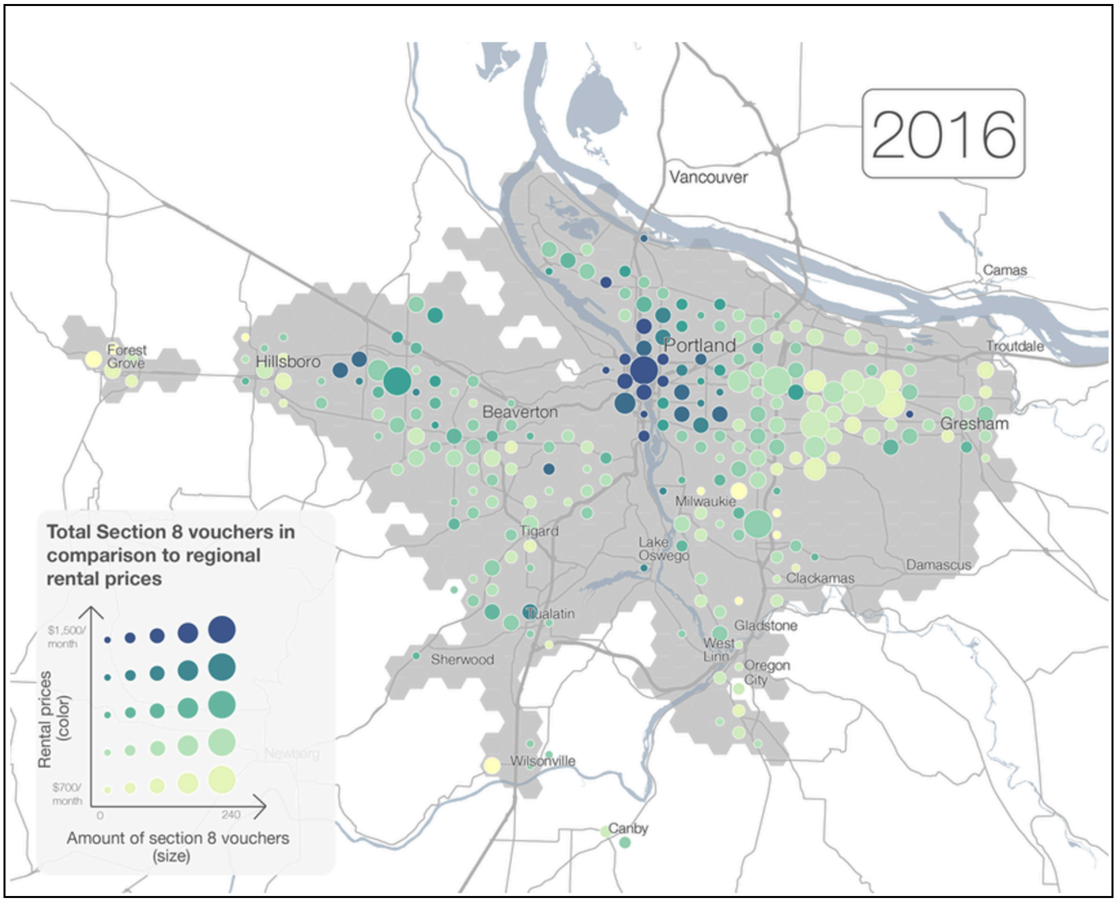


Figure 5.7. Proportional symbols (amount of Section 8 vouchers) with sequential colors to indicate a second variable (rental prices).

Credit: Oregon Metro, public domain

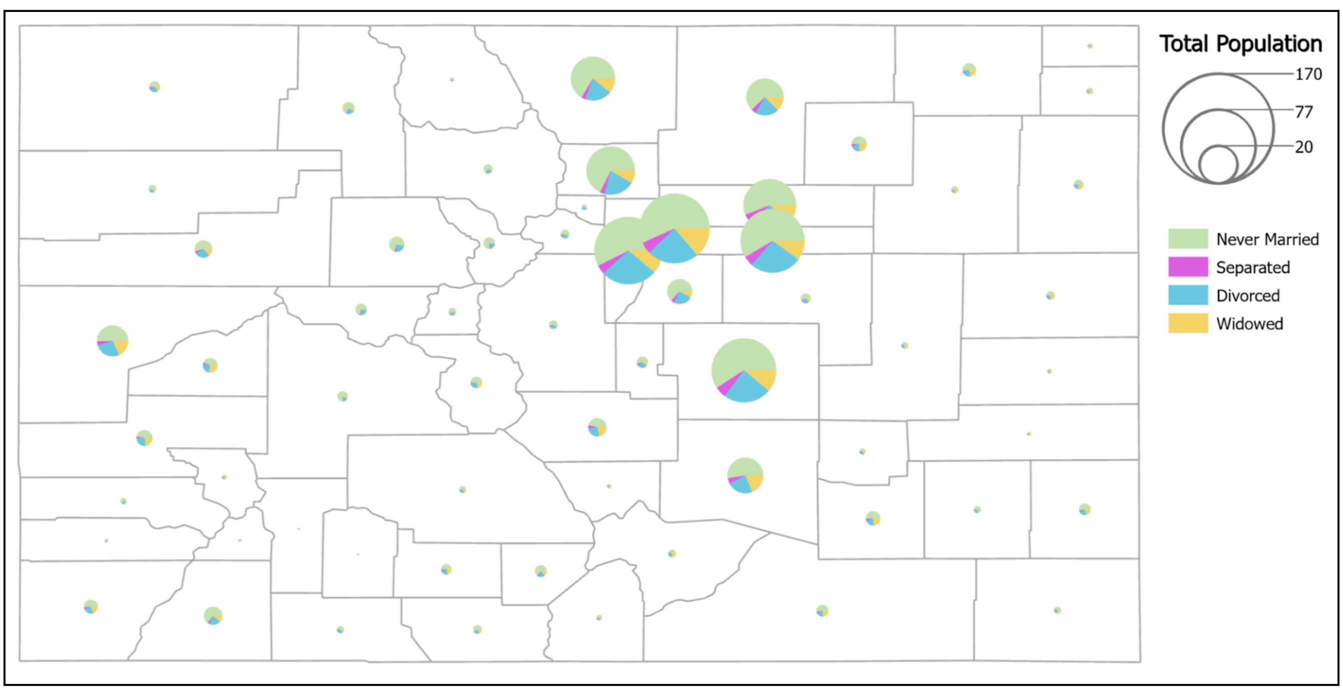


Figure 5.8. Colorado population map using proportional symbols for total counts and pie charts to show portions of the population based on marital status.

Data Source: US Census Bureau

Three-dimensional geometric symbols can also be used as proportional symbols as well. Examples of three-dimensional geometric symbols are a sphere, prism, or pyramid. The advantages of three-dimensional geometric symbols are that they are visually attractive and they allow for less crowding of the map. The disadvantage of three-dimensional geometric symbols is that readers cannot accurately judge scaling differences because they now have to judge changes in volume as well as area. Graduated symbols are recommended for three-dimensional representations to combat scaling issues [2].

Overlap: Overlapping symbols express a sense of visual cohesiveness and may make the mapped phenomena more perceptible. However, it can make it harder for the map reader to estimate the individual symbol sizes as they will be partially obscured, though using partially transparent symbols can help. Symbols should not overlap more than 25%, with the goal overlap being 10-15% (Figures 5.9 and 5.10) [2].

Figure 5.9. Both maps show an appropriate amount of overlap for the proportional symbols.
Credit: Appropriate Overlap on a Map, *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

Figure 5.10. Too little overlap, data pattern not evident (left); Too much overlap, some data and enumeration units obscured (right).

Credit: Little Overlap, Empty, Uninteresting (left) and Too Much Overlap (right), *Introduction to Cartography*, Ulrike Ingram,

Scaling Methods: For proportional symbols, there is absolute scaling and apparent magnitude scaling (also known as Flannery scaling). In absolute scaling, symbols scale directly proportionally to their data values and to each other. However, if truly proportional symbols are used (for every single value) the reader is not able to calculate the value for each circle, only differentiating a given number of symbol sizes effectively. Additionally, viewers underestimate large circle sizes relative to smaller circles (Ebbinghaus illusion). To combat these issues, apparent magnitude scaling can be utilized. It applies a factor to compensate for the user's underestimation of the area of the symbols. Typically this factor is used to increase the size of the circles faster than in absolute scaling (Figure 5.11) [2].

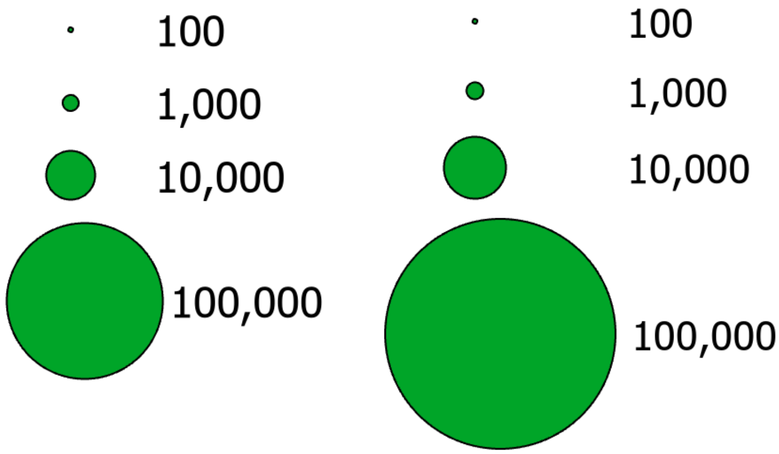


Figure 5.11. Absolute (left) and apparent magnitude (right) scaling for proportional symbols.

5.3 Graduated Symbols

Graduated scaling, also called range grading scaling, scales each symbol to represent a range of data values and not a single data value. Graduated symbols are similar to choropleth mapping as they employ data classification methods to determine class breaks. The main advantage of using graduated symbols is that readers can easily discriminate symbol sizes and match them to the legend symbols. Graduated scaling is recommended for three-dimensional symbols, such as prisms. As with proportional symbols, graduated symbols can be combined with other visual variables to create multivariate maps or be used on choropleth maps as shown in Figure 5.12 [2].

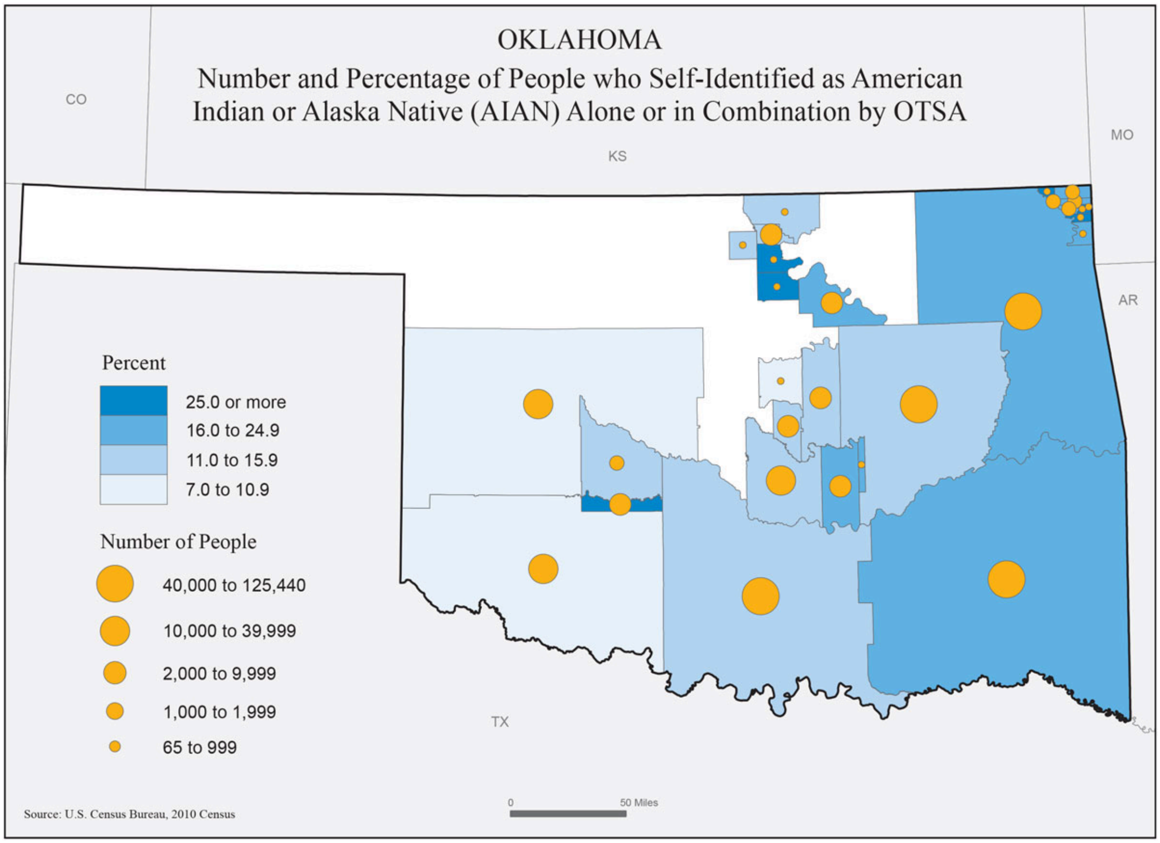


Figure 5.12. Graduated symbols on a choropleth map.
Credit: Static Proportional Symbol Map of Oklahoma, US Census Bureau, public domain

5.4 Legend Design

On proportional and graduated symbol maps, the legend serves as a visual anchor for interpreting symbol sizes. There are four common legend layouts, which are: vertical, horizontal, nested, and nested semi-symbols. In these different legend layouts, numbers or data values are placed to the right of the symbols for vertical, nested, and nested semi-symbol legends and below the symbols if it is a horizontal legend [2].

The main reason to use a nested symbols legend layout is that it requires less space on the map layout than the other legend options. To save even more space the values can be placed inside the symbols if there is space available, otherwise, the value should be to the right of the nested symbols with a leader line connecting the values to the symbols [2].

Vertical Legend: For the vertical legend layout (Figure 5.13), the values are displayed to the right of the symbol. Also, small values are at the top and large values are at the bottom

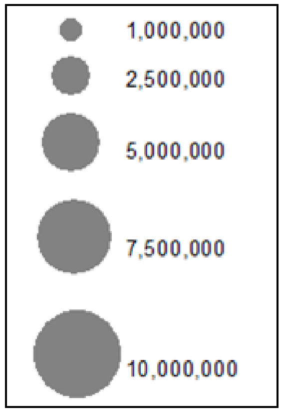


Figure 5.13. Vertical legend, proportional symbols.
Credit: Vertical Layout, adapted from *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

Horizontal Legend: For the horizontal legend (Figure 5.14), layout values are below the symbols, small values are on the left, and large values are on the right.

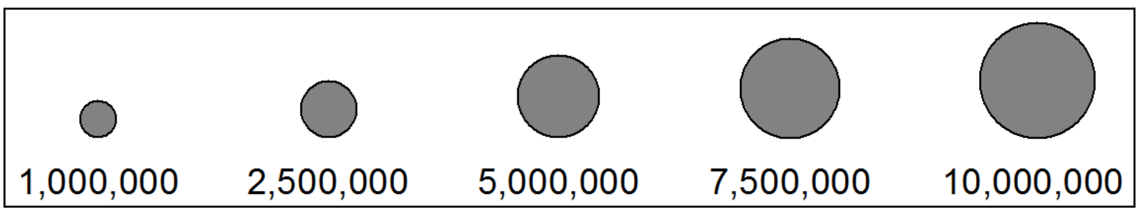


Figure 5.14. Horizontal legend, proportional symbols.
Credit: Horizontal Layout, *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

Nested Symbols Legend: In the nested symbols legend layout (Figure 5.15), the symbols are placed on top of each other with the smallest symbol on top and a line to the data value along the symbol tops or bottoms [2].

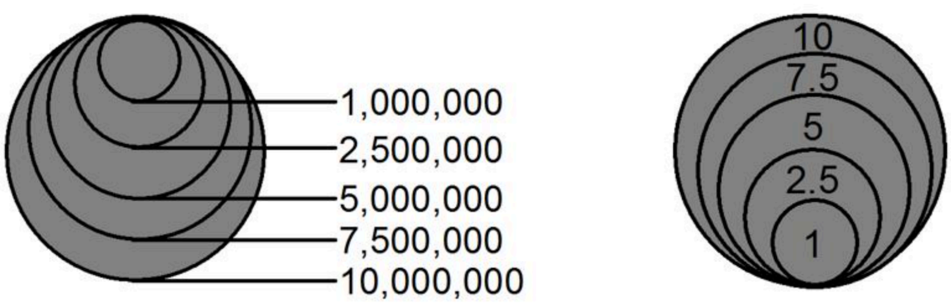


Figure 5.15. Nested symbol legend, proportional symbols.
Credit: Nested Symbol Layouts, *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

Nested Semi-Symbol Legend: The nested semi-symbols legend (Figure 5.16) layout requires the least amount of space on a map. They may have leader lines from the top or bottom of the symbols leading to the data values.

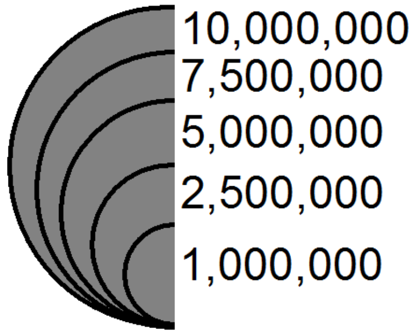


Figure 5.16. Nested, semi-symbol legend, proportional symbols.
Credit: Nested Semi-Symbol Layout, *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

5.5 Line Symbols

Line symbols are primarily used to symbolize linear features, such as roads and rivers. They can also be used to represent value as in a flow map. For quantitative variables, line size and separation (e.g., dashing) can reflect differences in order. For flow maps, lines can be proportional or graduated (same as with point symbols) in size to reflect the variable value. For qualitative variables, hue (Figure 5.17), shape, and orientation/angle are used to represent differences in kind [1].

When relating line representation to visual variables, wider lines represent higher values, more importance, higher intensity, etc., while separated/dashed lines represent lower values, less importance, etc. with the wider spacing between dashes indicating lower values. Line casing can be utilized by placing a line symbol on top of a wider line symbol, which functions similarly to a halo by increasing visual hierarchy [1].

For line placement, generalized locations are often used on flow maps such as shown in the bird migration figure below (Figure 5.17) or where only general location is known. For exact line placement is used for features in which the geographic location is known and important to the map function and purpose, such as a river basin map (Figure 5.18) [1].

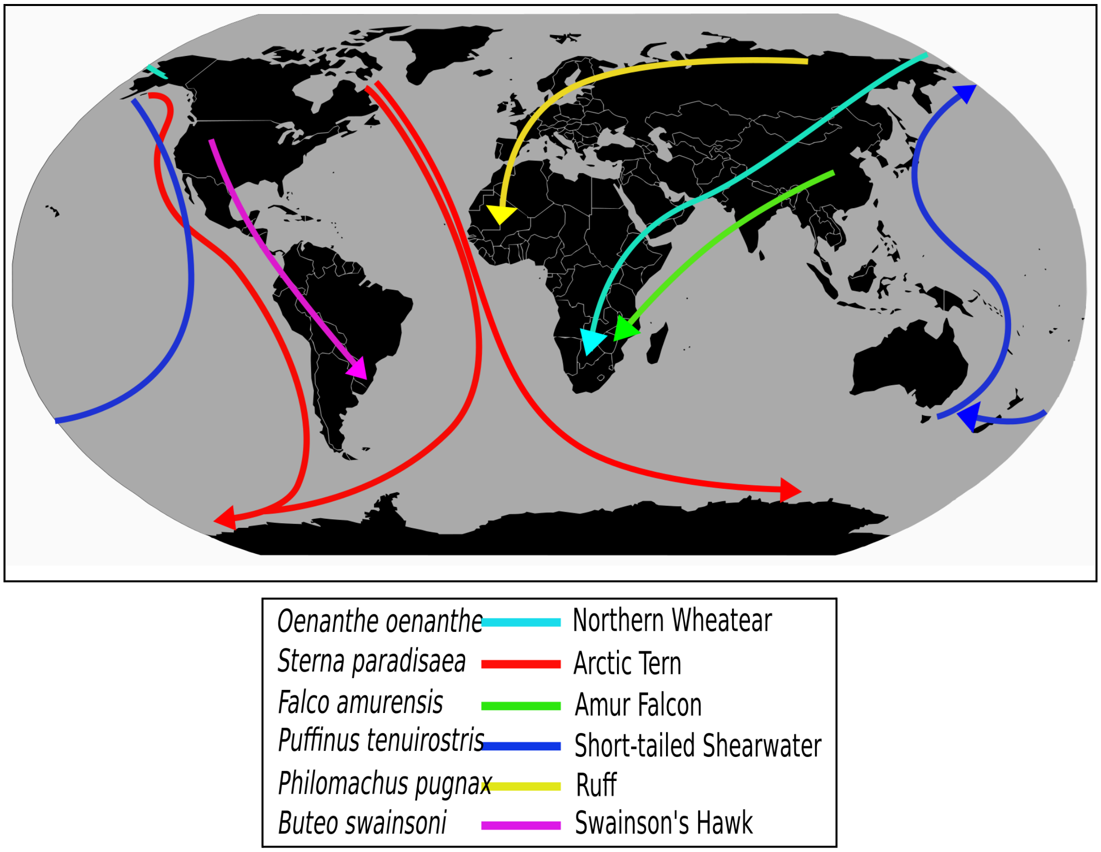


Figure 5.17. Selected long-distance bird migration routes are shown as a flow map. Hue differentiates the species.

Credit: adapted from Migrationroutes, [Wikipedia](), L. Shyamal, public domain

Figure 5.18. Colorado River Basin.
Credit: Colorado River Basin map, USGS, public domain

5.6 Area Symbols

Area symbols, also known as polygons in GIS, can represent continuous features such as land cover, or discontinuous features such as population. Any of the visual variables can be used to represent area data and their application varies based on whether the mapped data is qualitative (e.g., water resource regions - Figure 5.19) or quantitative (e.g., population growth - Figure 5.20). A special case for using size to map areas is the cartogram, where the quantitative variable is used to increase or decrease the size of a feature as shown in Figure 5.21, which focuses on the 2008 presidential election [1].

Figure 5.19. Qualitative map of the two-digit hydrologic unit codes (HUCs), of the United States of America, per the USGS hydrologic unit classification system.
Credit: Watershed Resource Regions, [USGS](), public domain

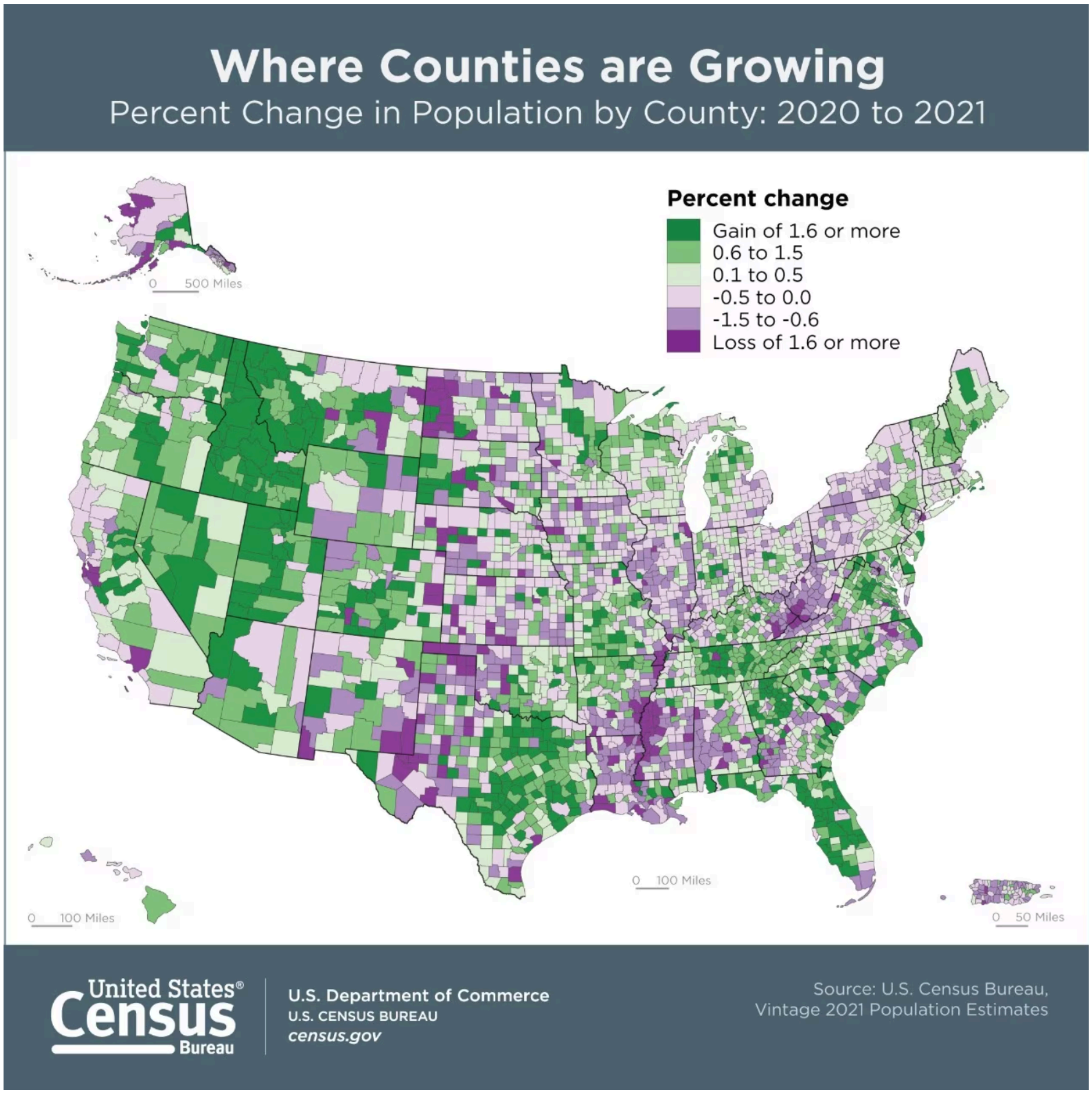


Figure 5.20. US County Population Change from 2020 to 2021.
Credit: adapted from "Where Counties are Growing", US Census Bureau, public domain

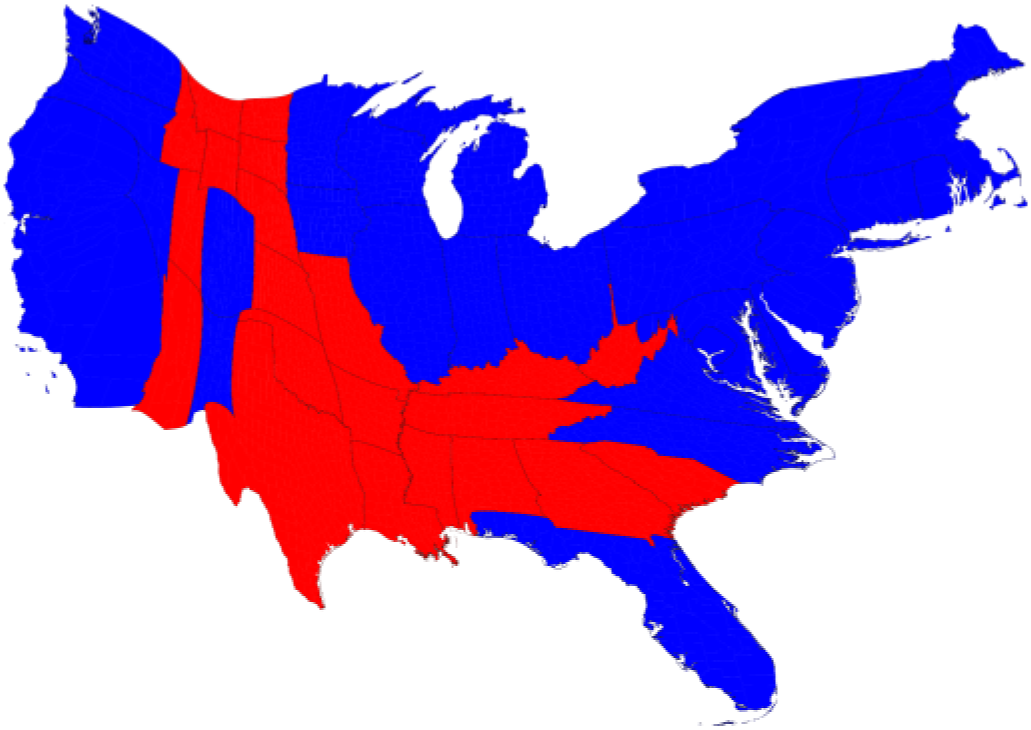


Figure 5.21. 2008 Presidential election cartogram (bivariate). The size of each state is scaled to be proportional to their number of electoral votes. Red states indicate the outcome was in favor of the republican candidate while blue indicates the outcome was in favor of the democratic candidate.
Credit: from "Maps of the 2008 US presidential election results", *University of Michigan*, Mark Newman, CC BY 2.0

When choosing symbols to represent areas, they can be literal such as the wetland symbol shown below, or abstract such as the wilderness area below that is represented solely by a hue of brown regardless of where the area is brown in reality (Figure 5.21). Literal patterns should represent the logical relationship of the data, such as using green trees for a forest not to represent a desert [1].

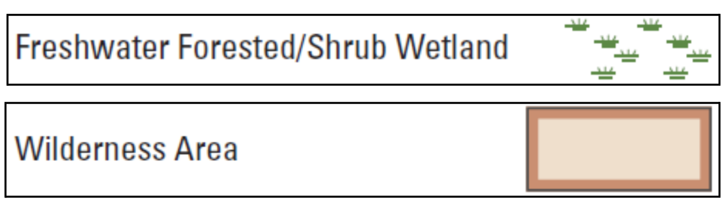


Figure 5.21. USGS legend symbols for freshwater forested wetlands and wilderness areas.
Credit: adapted from US Topo Map Symbol File Sample, USGS, public domain

Area data can be presented as an imaginary 3-D surface referred to as a statistical surface or isarithmic map. An isarithmic map uses lines, similar to elevation lines on contour maps, to represent areas of similar value, such as temperature. When color is applied in between lines to represent the areas of similar value it is called an isoplethic map as shown in Figure 5.22. Other common applications of isarithmic maps are those showing population or crime density [1].

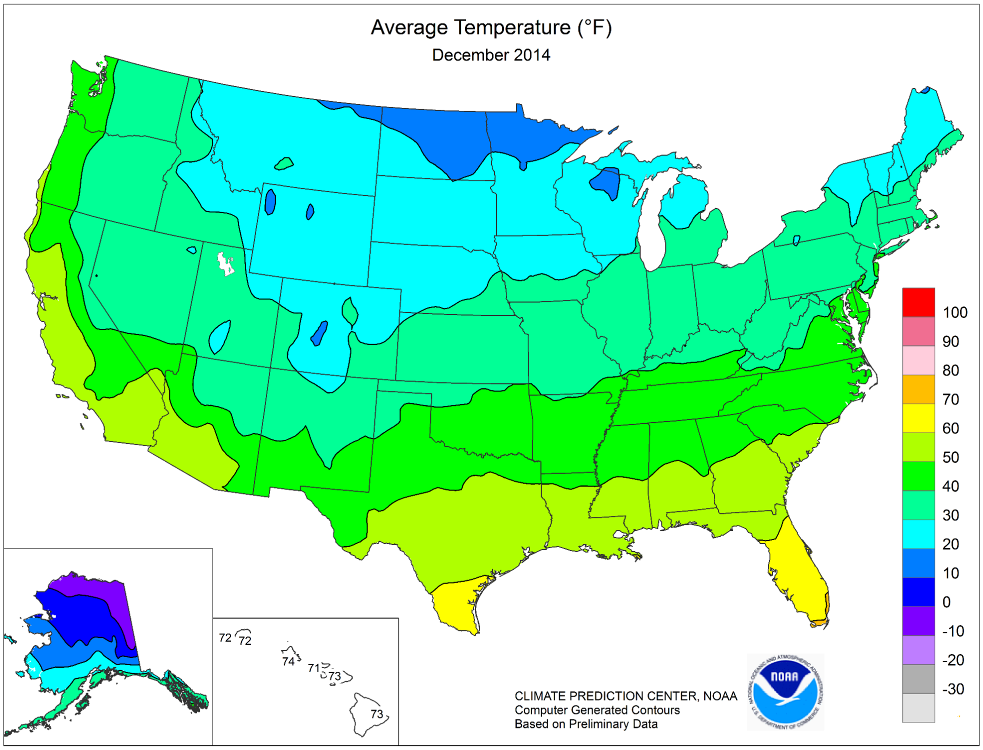


Figure 5.22. Isoline map of average temperatures in the United States for December 2014.
Credit: Average temperature, [NOAA, National Weather Service, Climate Prediction Center](), public domain

Additional Resources

Types of Thematic Maps - https://gistbok.ucgis.org/bok-topics/common-thematic-map-types [opens in new tab]

References - materials are adapted from the following sources:

[1] GEOG 3053 Cartographic Visualization by Barbara Buttenfield, University of Colorado Boulder, used with permission.

[2] Introduction to Cartography by Ulrike Ingram under a CC BY 4.0 license

Chapter 6
Color on Maps

For information on color theory and applying color schemes on maps see the following resources:

GIS&T Body of Knowledge: https://gistbok.ucgis.org/bok-topics/color-theory [opens in new tab]

Penn State University:
Color Overview - https://www.e-education.psu.edu/geog486/node/534 [opens in new tab]
Specifying Colors - https://www.e-education.psu.edu/geog486/node/606 [opens in new tab]
Type of Color Schemes - https://www.e-education.psu.edu/geog486/node/607 [opens in new tab]

NASA Earth Observatory:
Subtleties of Color, Part 1, Introduction - https://earthobservatory.nasa.gov/blogs/elegantfigures/2013/08/05/subtleties-of-color-part-1-of-6/ [opens in new tab]

Subtleties of Color, Part 2, The "Perfect" Palette - https://earthobservatory.nasa.gov/blogs/elegantfigures/2013/08/06/subtleties-of-color-part-2-of-6/ [opens in new tab]

Subtleties of Color, Part 3, Different Data Different Colors - https://earthobservatory.nasa.gov/blogs/elegantfigures/2013/08/12/subtleties-of-color-part-3-of-6/ [opens in new tab]

ColorBrewer2.0, Color Advice for Cartography: https://colorbrewer2.org/ [opens in new tab]

Chapter 7
Scale and Resolution

7.0 Representations of Map Scale

One of the most significant challenges behind mapping the world and its resident features, patterns, and processes is reducing it to a manageable size. What exactly is meant by “manageable” is open to discussion and largely depends on the purpose and needs of the map at hand. Nonetheless, all maps reduce or shrink the world and its geographic features of interest by some factor. Map scale refers to the factor of reduction of the earth’s features so they fit on a map [3]. A map’s scale can be represented in three ways - a verbal scale, a graphic scale or scale bar, or a representative fraction (Figure 7.0). In all scale representations, the purpose is to demonstrate the relationship between the size/distance represented by the map features versus their size/distance on the earth’s surface, in other words, the map distance relative to the ground distance. Scale can also be defined as the amount of size reduction of the map features.

Graphic Scale / Scale Bar: 

Verbal Scale: One inch on the map equals twenty feet on the ground

Representative Fraction: 1:24,000

Figure 7.0. The three types of scale used on maps.
Credit: Graphic Scale, *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0 (graphic scale only)

Verbal scale: A verbal scale uses words to describe the relationship between the map distance and ground distance. For example, “one inch equals one mile” gives the map reader a useful explanation of the map scale.

Graphic Scale / Scale Bar: A graphic scale or scale bar is a graphical representation of distance on a map. Scale bars are often more useful than the other representations of scale when the map reader needs to measure to estimate distances on a map. One important advantage of scale bars is that they remain true when maps are shrunk or magnified, unlike the other two types of map scale [4]. Scale bars should be easy to read, not overly large or small relative to the map feature, not be overly complicated or ornate, and use round numbers (100 vs. 105.5). An important concept to keep in mind is that the scale bar is the only scale format that enlarges and reduces relative to the map itself, therefore should be used in situations where scale needs to be represented, but the original map size may be modified.

Representative Fraction: The representative fraction (RF) describes scale as a simple ratio, also called a ratio scale. The numerator, which is always set to one (i.e., 1), denotes map distance and the denominator denotes ground or “real-world” distance. One of the benefits of using a representative fraction to describe scale is that it is unitless. In other words, any unit of measure can be used to interpret the map scale. Consider a map with an RF of 1:10,000. This means that one unit on the map represents 10,000 units on the ground. Such units could be inches, centimeters, or any unit of measurement selected by the map reader [3].

7.1 Large Scale and Small Scale Maps

Map scales can also be described as either “small” or “large.” Such descriptions are usually made based on the representative fractions and the amount of detail represented on a map. For instance, a map with an RF of 1:1,000 is considered a large-scale map when compared to a map with an RF of 1:1,000,000 (i.e., 1:1,000 > 1:1,000,000). In other words, the smaller the denominator the larger the scale (Figure 7.1). Furthermore, while the large-scale maps show more detail and less area, the small-scale maps show more area but less detail [3] (Figure 7.2).

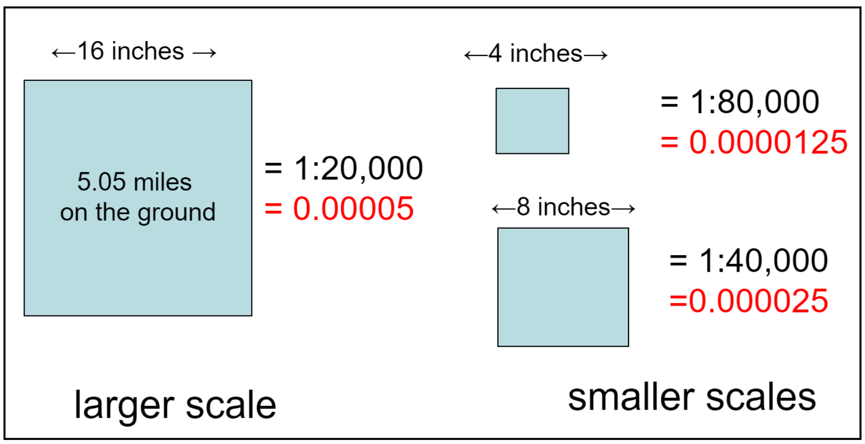


Figure 7.1. Large vs. small scale based on the representative fraction
Credit: graphic by Barbara Buttenfield, University of Colorado Boulder, used with permission

Generally, maps with an RF larger than 1:50,000 (e.g., 1:15,000) are considered large-scale maps, maps with an RF of 1:50,000 – 1:250,000 are of intermediate scale, while maps with an RF smaller than 1:250,000 (e.g., 1:1,000,000) are considered small scale [1].

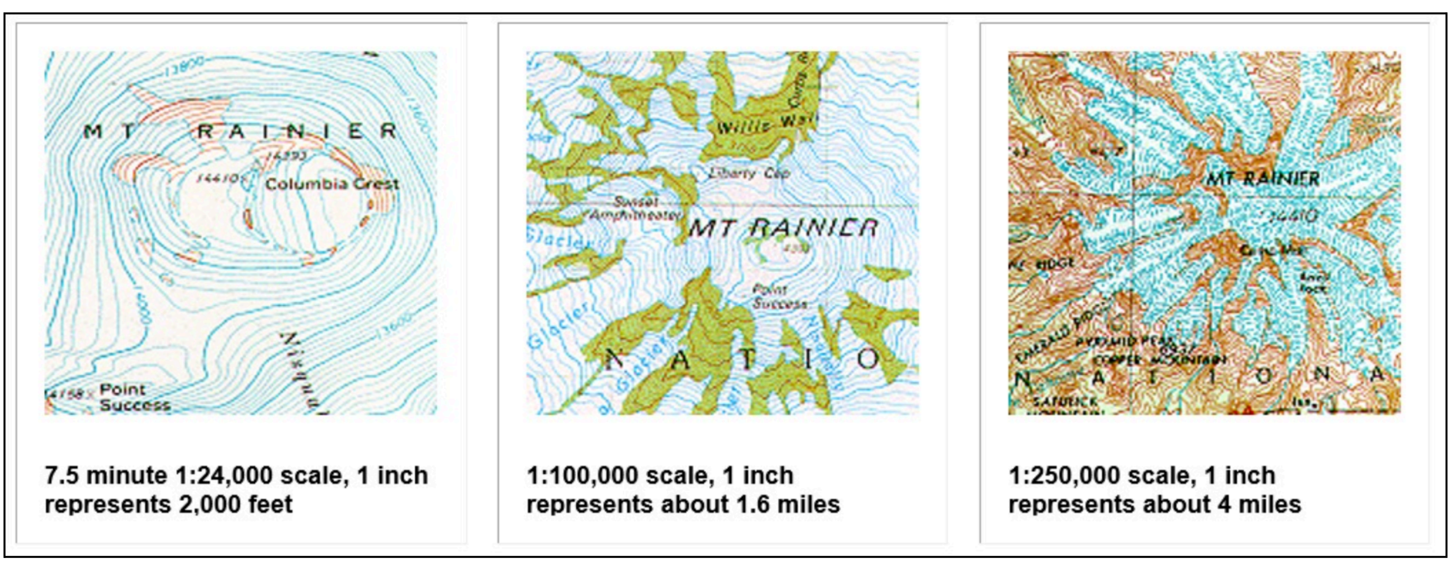


Figure 7.2. Large, intermediate, and small-scale maps of Mt. Rainier in Washington, USA.
Credit: Common Mapping Scales, USGS, public domain

Choosing an Appropriate Scale: Scale choices, first and foremost, should be based on map purpose and expected audience. It will also vary based on the format in which the map will be accessed (e.g., paper map, trailhead sign, or online) and whether or not it is an interactive, web-based map or a static map. The resolution of the data being used will also guide the cartographer towards an appropriate scale. Experienced map makers will ask the following questions before deciding on a map scale - a) what area (geographic footprint) can be shown?; and b) what level of detail can be shown within that footprint? There will always be a tradeoff between the necessary details and the available space in which to fit those map details so it is up to the map designer to make an informed decision based on the factors indicated above [1].

All maps possess a scale, whether it is formally expressed or not. Though some say that online maps and GIS are "scaleless" because we can zoom in and out at will, it is probably more accurate to say that GIS and related mapping technology are multiscalar. Understanding map scale and its overall impact on how the earth and its features are represented is a critical part of both map making and GIS [3].

Representing Scale on Maps: Scale does not always need to be represented on a map, nor in all three formats. Many thematic maps will not require a scale at all, such as that of an election map of the United States, since for such maps, the map reader does not need to measure distance explicitly. One expectation would be that if the map is prepared for users not familiar with the United States; here it is only applied to give the map reader an understanding of the overall scale and the size difference between states. All reference maps should have at least one scale representation, typically a scale bar, especially if the map will be used for navigation or distances need to be measured. World maps almost always include the representative fraction. Topographic maps produced by the USGS always have all three scale formats present.

7.2 Resolution

This section focuses on spatial resolution and how to choose an appropriate resolution based on the map scale. Spatial resolution describes the smallest unit that is mapped [3] or the size of the pixel when referring to raster data. It can also refer to the smallest detectible object on a map or image or the level of detail that can be seen.

When discussing rasters it is best to refer to it as either coarse or fine resolution, rather than low or high, though both are used in the GIS community. In Figure 7.3, the image shows a vector feature in the first column, then the same feature represented as a 30-meter resolution raster. Note how the representation of the feature changes; the area represented by each feature changes as well.

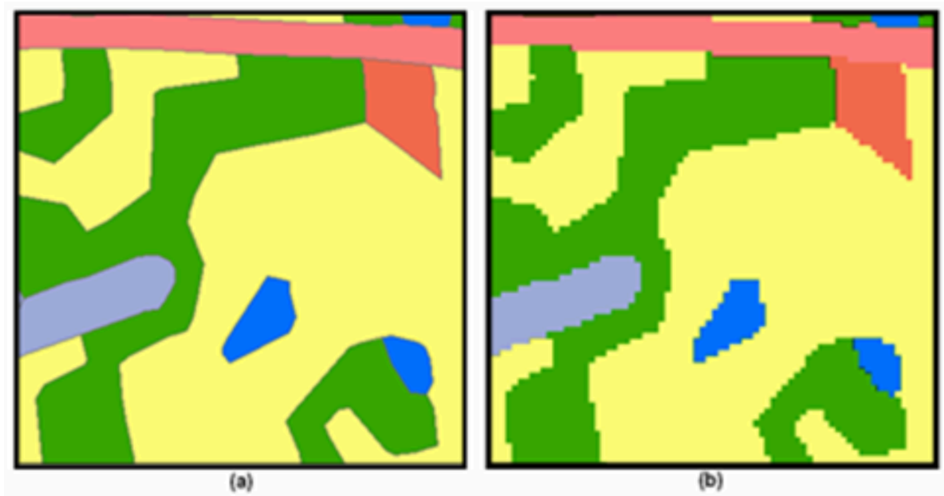


Figure 7.3. Comparing a vector feature (a) and its representation in raster format (b) with 30m resolution.

Credit: Comparison of land-use and land-cover data, adapted from USGS, public domain

Determining Appropriate Spatial Resolution: The spatial resolution refers to the dimensions representing the area of the ground covered by each cell. The finer the resolution, the smaller the cell size, and therefore the more detail the raster will have. As shown in Figure 7.4, the detectable size of a feature is twice that of the raster resolution. For example, if a feature is 10 meters in size, the raster resolution (cell size) needs to be 5 meters (5x5 meter cell) in order to detect the feature in the raster image. Figure 7.4 also indicates the largest scale at which a particular raster resolution should be used. The spatial resolution of a raster needed will also vary based on the particular project, including data availability, what type of map is being created, what type of analysis is being performed, and the storage space available for such data [5].

Map Scale	Detectable Size (in meters	Raster Resoltuion (in meters)
1:1,000	1	0.5
1:10,000	10	5
1:25,000	25	12.5
1:100,000	100	50
1:250,000	250	125
1:1,000,000	1,000	500

Figure 7.4. Relationship between map scale, feature detection, and raster resolution.

Credit: Sarah Schlosser, University of Colorado Boulder

Examine the visual difference between the two images below (Figure 7.5). Note that the resolution is the same, but the scales are different thereby changing what can be detected in the images.

Figure 7.5. Effect of scale changes with constant resolution.
Left 1:145,000, right 1:18,000 both using NAIP Plus 1-meter resolution data.
Data Source: *USGS, The National Map*

Now examine the differences between raster resolutions with a constant scale (Figure 7.6). Note what can be detected and differentiated in the image decreases with increasingly coarser resolution data.

Figure 7.6. Images at the same scale with different raster resolutions, coarse (left) and fine (right).
Credit: Medium Resolution Compared to High Resolution Image of USGS EROS, USGS, public domain

Converting Between Data Resolution and Scale: Dr. Waldo Tobler, a renowned cartographer, stated that, "the rule is: divide the denominator of the map scale by 1,000 to get the detectable size in meters. The resolution is one-half of this amount" [5].

For example, if you wanted to determine what resolution imagery is needed to detect features at a map scale of 1:10,000, using Tobler's rule, imagery of approximately 5-meter resolution would be effective [10000/ (1000*2)] [5].

To determine the appropriate scale for a particular raster dataset, the following formula can be applied:

$$\text{Pixel Size} * 1000 \sim \text{RF denominator}$$

For example, if you have a raster dataset with a resolution (pixel size) of 30m, the appropriate scale for use is approximately 1:30,000. The raster resolution needs to be expressed in meters when using this formula. This formula provides an estimate and generally, data can be used in map scales up to 20% larger than yielded by the formula. For our example map scale of 1:30,000, the cartographer could use the 30m data in a map with a scale as large as 1:24,000 without significant consequence. In going from a larger to a smaller scale, you can generalize (modify details) the raster to decrease the resolution. However, when going from smaller to larger scales it is rare to jump to a larger scale, beyond a 20% increase. In that situation finer resolution data would need to be used [1].

Additional Resources

Resolution: https://gistbok.ucgis.org/bok-topics/resolution [opens in new tab]
Scale: https://gistbok.ucgis.org/bok-topics/scale-and-generalization-1 [opens in new tab]

References - materials are adapted from the following sources:

[1] GEOG 3053 Cartographic Visualization by Barbara Buttenfield, University of Colorado Boulder, used with permission.

[2] Introduction to Cartography by Ulrike Ingram under a CC BY 4.0 license

[3] Introduction to Geographic Information Systems by R. Adam Dastrup under a CC BY 4.0 license.

[4] Mapping, Society, and Technology by Steven M. Manson under a CC BY-NC 4.0 license

[5] On map scale and raster resolution by Rajinder Nagi, Esri, 2010.
https://www.esri.com/arcgis-blog/products/product/imagery/on-map-scale-and-raster-resolution/

Chapter 8
Generalization

8.0 Generalization

Generalization refers to the selection, simplification, and symbolization of detail according to the purpose and scale of the map. Mapmakers should approach the task of generalization with caution, ensuring they fully understand both the purpose of the map and the characteristics of the items to be generalized before proceeding. The overall goal is to preserve the geographic patterns present in the information while still meeting the map goals at the desired scale and improving communication and legibility of the map [2].

8.1 Goals of Generalization

As it is not possible to communicate map information at a 1:1 scale, generalization has many goals, including the following:

Structure:
- The map content is well structured.
- The map content priorities are adapted to the map's scale and the intended purpose.
- The objects are classified according to clear and reasonable criteria.
- The grouping of objects is logical.

Legend:
- Expressive and associative symbols constitute the base for clear map communication.
- The size and the form of the symbols are adapted to the other symbols and to the reality.

Generalization level:
- The level/amount of generalization varies according to the purpose and the scale.
- The level of generalization is carefully defined.

Selection of objects:
- The object selection complies with the map's scale and with the intended purpose.
- The objects that are visible in reality (e.g., houses) are indicated with non-visible objects such as borders or labelling.

Accuracy of objects:
- The optimal accuracy of the objects regarding position and form is reached.
- Displacing objects is only needed for increasing the legibility and for differentiating map features.
- The symbols of visible objects (in reality) have a high accuracy.

Reality accuracy:
- While the reality is revised and changed, the map features represent truth.

- All objects present in the map really exist.
- Appropriate legend symbols are assigned to the objects.
- Labelling is correctly written and assigned.

Legibility of the map elements:
- Good legibility is achieved with respect to the graphical minimal dimensions (sizes and distances) of the symbols.
- Graphical readability rules support legibility.
- Legend is credible and exact.
- The quantitative generalization of objects (e.g., houses) respects the density of objects in reality.
- The relations and dependencies of objects in reality (e.g., streets, rivers, contour lines, etc.) are carefully considered.

Credit: Aims of Generalization, GITTA, Stern et al., CC BY-NC-SA 2.5 [1]

8.2 Types of Generalization

Selection and Omission: Selection is performed by the mapmaker to determine what to include and omission is what they determine to exclude from the map. No modification of the data is required during either process. For example, the mapmaker can decide to depict secondary roads or only major highways, to include only cities with a population of 50,000 people or more, or only include perennial streams (selecting perennial streams and omitting intermittent streams).

Simplification: Simplification is the process which determines the important characteristics of the data, the retention of these important characteristics, and the elimination of the unwanted details (Figures 8.0 and 8.1). When reducing the scale of a map, each map item will occupy a proportionately larger amount of space. Consequently, simplification must be practiced in order to ensure legibility and truthful portrayal. The elimination of unwanted map element details (points or features) is the most often used form of simplification. Below are two figures showing examples of simplification [1].

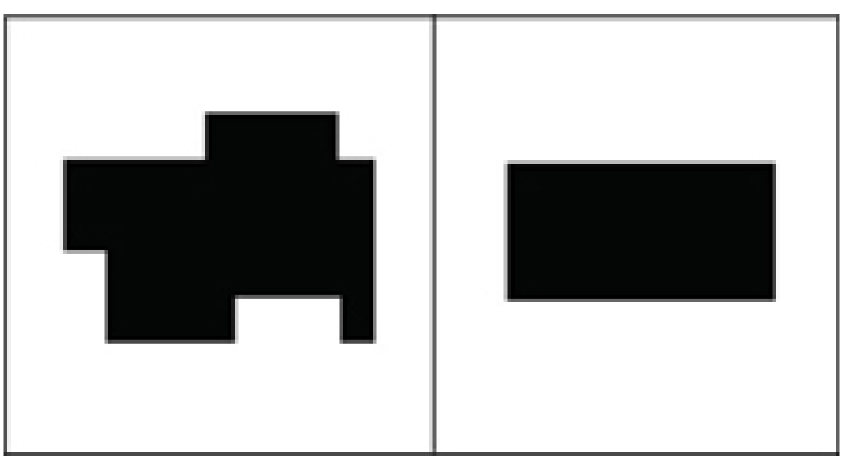


Figure 8.0. Area simplification of a building.
Credit: Shape simplification, GITTA, Stern et al., CC BY-NC-SA 2.5

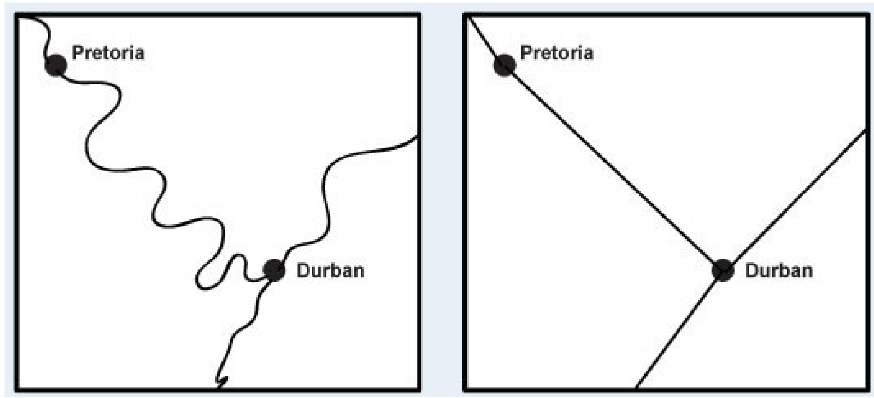


Figure 8.1. Line feature simplification.
Credit: Image from [GITTA,](#) Stern et al., [CC BY-NC-SA 2.5](#)

Smoothing: Smoothing is the adjustment of the geometry of features on a map in order to maintain a realistic appearance (Figure 8.2). Shorelines, rivers, and borders between countries often have lots of curves and bends. When working at small scales, a mapmaker typically chooses to first simplify and then smooth the shapes of objects and lines [3].

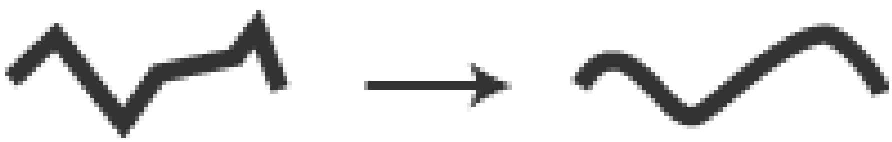


Figure 8.2. Line smoothing.
Credit: Smoothing: reduce sharp angles to smoother curves, *Cartography Guide*, [Axis Maps](#), [CC BY-NC-SA 4.0](#)

Grouping / Aggregation: Aggregation, sometimes called grouping, is the process in which information is gathered and expressed in a summary form (Figure 8.3). As with classification, a common aggregation purpose is to bring relative order and simplicity [1]. For example, a wind farm can be expressed as an area (aggregation) instead of showing individual wind turbines or county data can be aggregated to the state level.

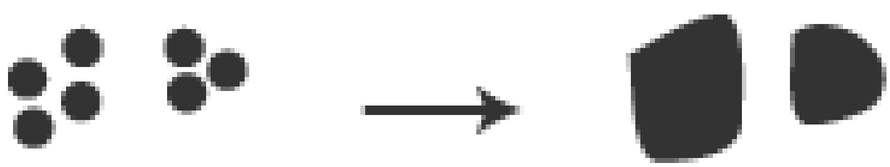


Figure 8.3. Feature aggregation - points to areas.
Credit: Aggregation: group points into an area, *Cartography Guide*, [Axis Maps](#), [CC BY-NC-SA 4.0](#)

Displacement / Exaggeration: When map elements are too close together after scaling, they must be moved in order to avoid their visual merging through displacement or exaggeration (Figures 8.4 and 8.5). This often occurs when features of the same type are in close proximity. Exaggeration specifically aims to make an element seem larger or more important, than they are in reality [1]. Both exaggeration and displacement are used to increase the visibility and legibility of the features.

Figure 8.4. Feature displacement.
Credit: Displacement: separate objects (for clarity), *Cartography Guide*, [Axis Maps](), [CC BY-NC-SA 4.0]()

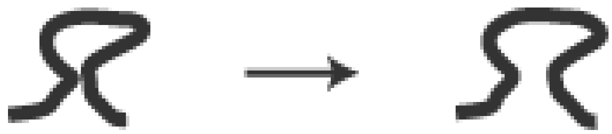


Figure 8.5. Feature exaggeration.
Credit: Exaggeration: amplify a part of an object (for clarity), *Cartography Guide*, [Axis Maps](), [CC BY-NC-SA 4.0]()

Classification and Symbolization: Anytime we classify or categorize data, generalization is being performed since individual values are lost. Classification is done in order to clarify and emphasize patterns of distribution of a particular phenomenon. See Chapter 10 for more information on data classification. Symbolization is the selection and design of symbols to represent geographic phenomena on a map. Many cartographers feel it should not be included with generalization and for the purposes of this text is discussed in Chapter 5.

Additional Resources
Scale and Generalization: https://gistbok.ucgis.org/bok-topics/scale-and-generalization-1 [open in new tab]
Feature Generalization: https://gistbok.ucgis.org/bok-topics/point-line-and-area-generalization [opens in new tab]

References - materials are adapted from the following sources:

[1] Generalisation of Map Data by Boris Stern, Marion Werner, Lorenz Hurni, and Samuel Wiesmann at the Geographic Information Technology Training Alliance under a CC BY-NC-SA 2.5 license

[2] GEOG 3053 Cartographic Visualization by Barbara Buttenfield, University of Colorado Boulder, used with permission.

[3] Mapping, Society, and Technology by Steven M. Manson under a CC BY-NC 4.0 license

Chapter 9
Datums, Coordinate Systems, and Map Projections

9.0 What is Geodesy?

Geodesy is the scientific discipline that deals with the measurement representation of the Earth. Geodesists study the Earth's gravitational field, its geometry, the motion of the Earth's crust, tides, and the Earth's rotation.

The Earth's Shape: There are three commonly used approximations of the shape of the Earth: sphere, ellipsoid, and geoid. The sphere is the simplest approximation of the Earth's shape as it is defined by a single radius with all points on the sphere's surface being equidistant from the center point. The ellipsoid is closer to the Earth's shape as defined by semi-major and semi-minor axes, with one axis being slightly longer than the other. The geoid is the closest approximation of the Earth's shape based on gravity measurements (Figure 9.0) [4]. It is an uneven surface that approximates the mean sea level across the Earth's surface by taking gravitational forces into account [2]. The sphere and ellipsoid are still used most often in mapping as they are much easier to apply to projection calculations and, when defined properly, can be exceptionally accurate [4].

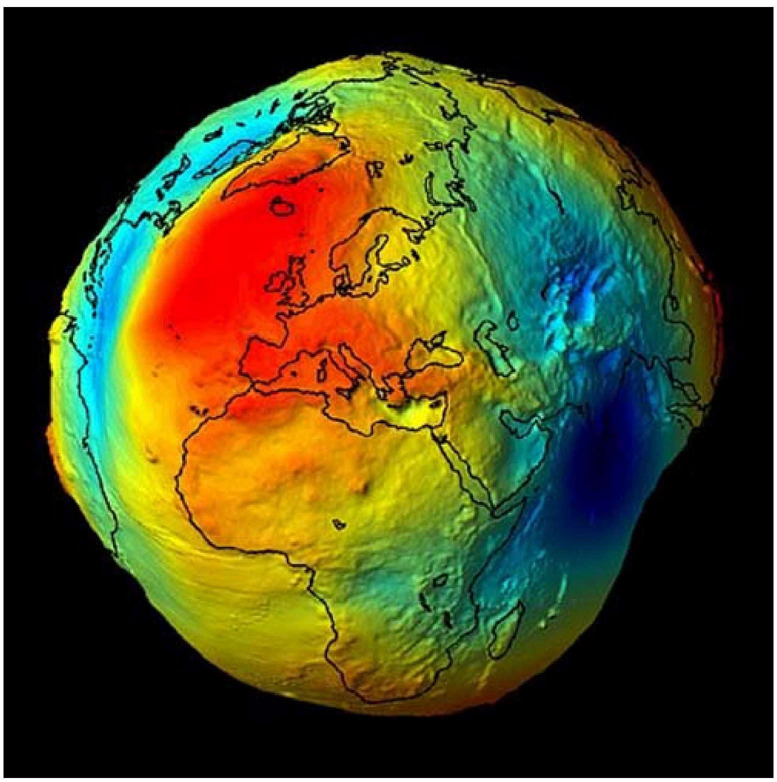


Figure 9.0. Exaggartation of the Earth's Geoid.
Credit: The Earth's gravity field (geoid), European Space Agency, public domain

Ellipsoids: Due to the centrifugal force created by the Earth's rotation, the Earth bulges slightly at the center—it is wider around the equator than from pole to pole. Because of this, a better way to describe Earth's shape is as an ellipsoid. Ellipsoids that closely resemble spheres are often called spheroids (Figure 9.1) [2].

The ellipsoid is the most commonly used of the three approximations of the Earth's shape today. There is not a single ellipsoid that fits Earth well; therefore, there are many official ellipsoids. Different ellipsoids were adopted because different measurements were used in different countries when defining these ellipsoids. Therefore, an ellipsoid created by one country may approximate that portion of the Earth very well, but may not be a suitable approximation for another portion of the Earth [4].

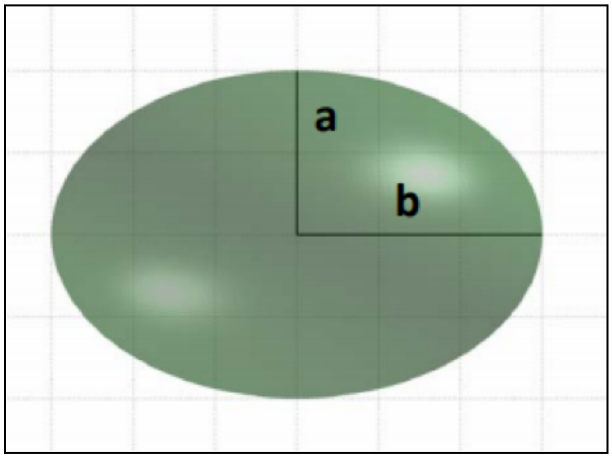


Figure 9.1. An exaggerated example of a spheroid.
Credit: oblate spheroid, [USGS](), public domain

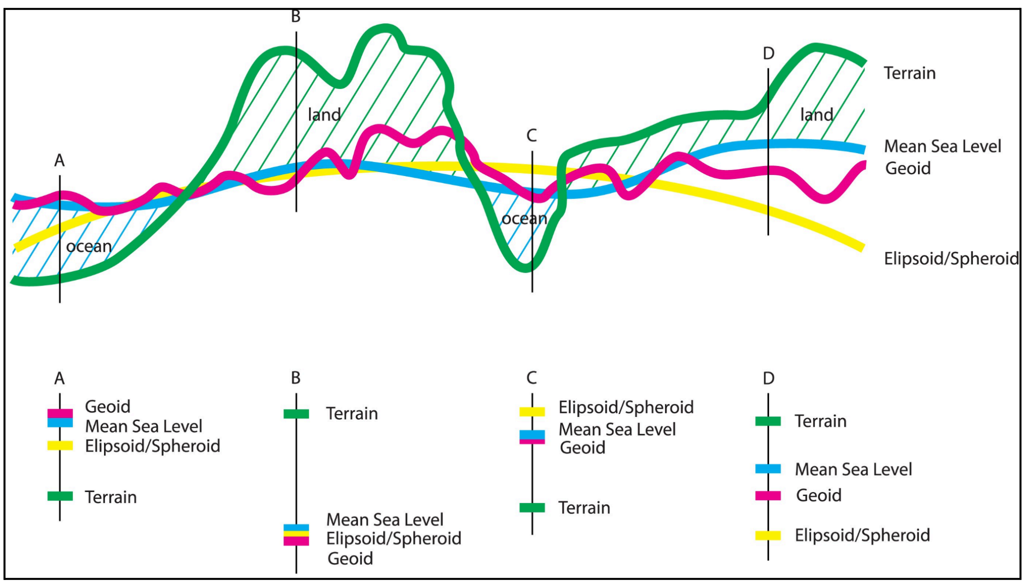


Figure 9.2. Comparison of the Earth's mean sea level, surface terrain, geoid and ellipsoid/spheroid.
Credit Image from [Intergovernmental Committee on Surveying and Mapping, Australia](), [CC BY 3.0]()

9.1 Datums

There are three common datums used in North America: North American datum of 1927 (NAD 27), North America datum of 1983 (NAD 83), and World Geodetic System of 1984 (WGS 84).

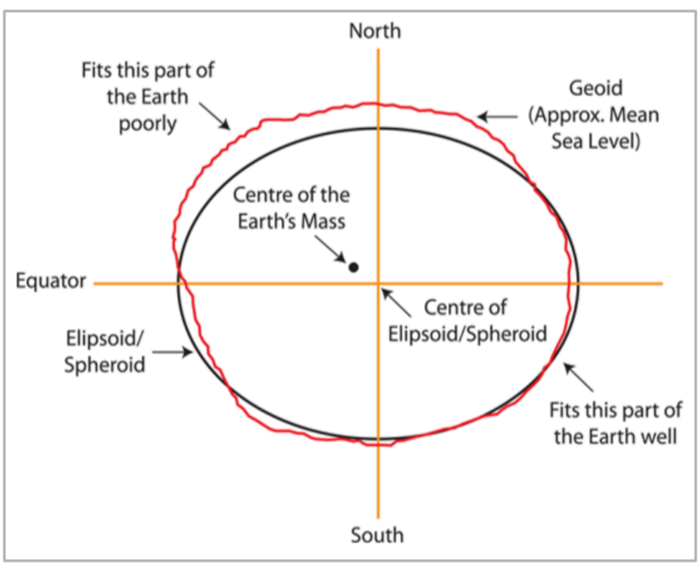
According to the National Geodetic Survey https://geodesy.noaa.gov/ [oepns in new tab], a datum is a set of constants specifying the coordinate system used for calculating coordinates of points on Earth

[4]. A datum specifies precisely the orientation and origins of the lines of latitude and longitude relative to the center of the Earth, a spheroid, or an ellipsoid [3]. We cannot assign any coordinates to a location without first specifying a datum and linking that datum to the shape of the Earth through field measurements [4].

NAD 27: The North American Datum of 1927 (NAD 27) is based on the Clarke 1866 ellipsoid which holds a fixed latitude and longitude in Meade's Ranch, Kansas. The locations were adjusted based on about 26,000 measurements across North America.

NAD 83: The North American Datum of 1980 (NAD83) is the modernized replacement of NAD27 and sought to improve positional accuracy as a result of adding thousands of new benchmarks [2]. It uses an Earth-centered reference ellipsoid rather than a fixed station (Figure 9.3). Additionally, 250,000 points were measured to adjust the latitude and longitude locations for improved accuracy [4].

WGS 84: The World Geodetic System of 1984 (WGS 84) is based on satellite measurements and the WGS 84 ellipsoid [4]. Because the datum uses the Earth's center as its origin, locational measurements tend to be more consistent regardless of where they are obtained. However, they may be less accurate than those returned by a local datum. Switching between datums will alter the coordinates (i.e., latitude and longitude) for all locations of interest and a datum transformation may be needed [3]. It is also important to note that the WGS 84 datum is used by Global Positioning Systems (GPS) to report latitudes and longitudes [4].



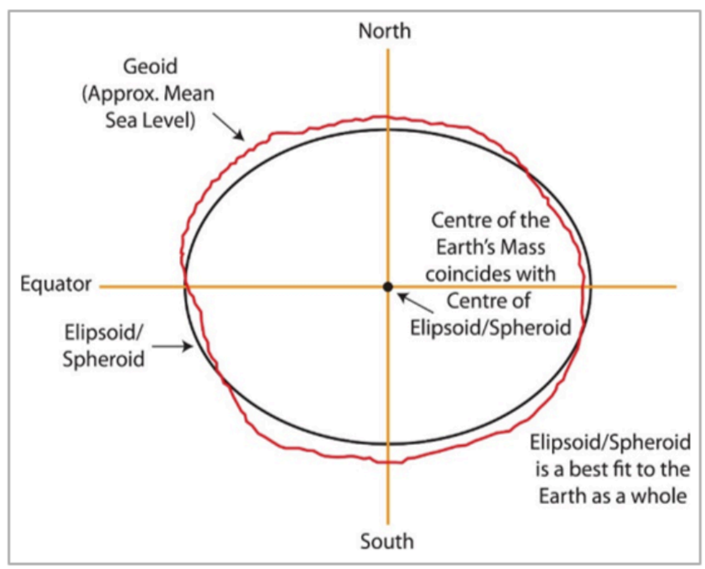


Figure 9.3. Australian Geodetic Datum (AGD84) (left), and the Geocentric Datum of Australia (GDA94) (right).

Credit: Image from Intergovernmental Committee on Surveying and Mapping, Australia, CC BY 3.0

Datum Transformation: Datum transformation is the process of re-calculating locations based on a different datum and may be necessary if you are combining datasets that were specified using different datums (e.g., NAD27 vs. NAD83), or if you would liketo map historical data using a more up-to-date system [4].

9.2 Coordinate Systems

Geographic Coordinate System (GCS): Coordinate systems are frameworks that are used to define unique positions. The coordinate system that is most commonly used to define locations on the three-dimensional Earth is called the geographic coordinate system (GCS), and it is based on a sphere or spheroid. A spheroid is simply a sphere that is slightly wider than it is tall and approximates more closely the true shape of the Earth [1].

The unit of measurement in the GCS is degrees, and locations are defined by their respective latitude and longitude within the GCS. Latitude is measured relative to the equator at zero degrees, with maxima of either ninety degrees north at the North Pole or ninety degrees south at the South Pole. Longitude is measured relative to the prime meridian at zero degrees, with maxima of 180 degrees west or 180 degrees east [1].

Note that latitude and longitude can be expressed in degrees-minutes-seconds (DMS) or in decimal degrees (DD) (Figure 9.4). When using decimal degrees, latitudes above the equator and longitudes east of the prime meridian are positive, and latitudes below the equator and longitudes west of the prime meridian are negative [1].

Nominal location	Absolute location (DMS)	Absolute location (DD)
Los Angeles, US	34° 3′ North, 118° 15′ West	+34.05, –118.25
Mumbai, India	18° 58′ North, 72° 49′ East	+18.975, +72.8258
Sydney, Australia	33° 51′ South, 151° 12′ East	–33.859, 151.211
Sao Paolo, Brazil	23° 33′ South, 46° 38′ West	–23.550, –46.634

Table 9.4. Example geographic coordinates in DMS and DD.
Credit: Table from *Essentials of Geographic Information Systems*, Campbell and Shin, CC BY-NC-SA 3.0

Longitude: Longitude is the angle of rotation measured east and west around the globe with the lines of longitude, also known as Meridians, running north-south from the North Pole to the South Pole. Lines of longitude will vary from positive 180° east to -180° west measured relative to the line of longitude of 0° which is known as the Prime Meridian, which runs through Greenwich, England. Lines of longitude west of the Prime Meridian up to and including 180° are represented as a negative number or as a Western longitude. Lines of longitude east of the Prime Meridian up to, and including 180°, are represented as a positive number or as an Eastern longitude [4].

Latitude: The second angle of rotation is known as latitude and the lines of latitude are referred to as parallels. Parallels measure north to south on the globe, and the lines run in parallel to each other from the North Pole to the South Pole. The equator is the latitude of 0°. Lines of latitude measure from positive 90° north, which is located at the North Pole, to -90° south, which is located at the South Pole [4].

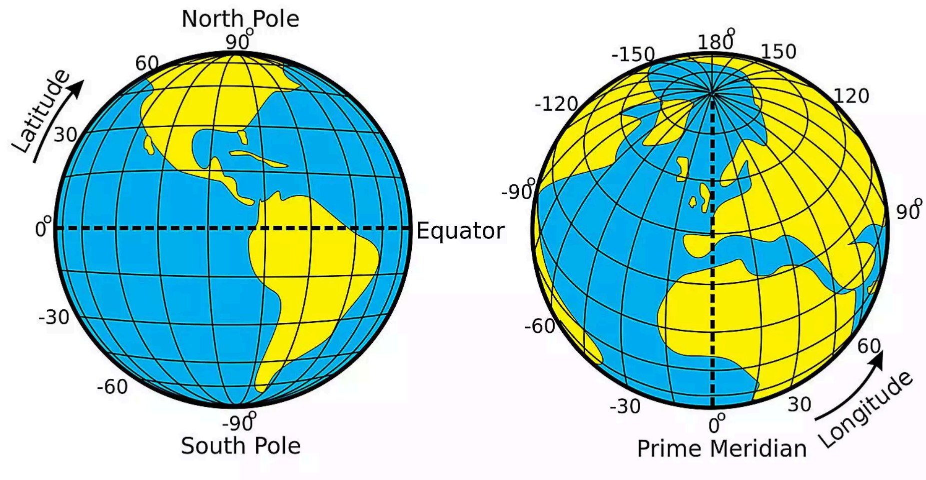


Figure 9.5 Latitude and Longitude.
Credit: Latitude and Longitude of the Earth, *Wikimedia Commons*, Djexplo, CC0 1.0

2D Coordinate Systems: A 2-D Cartesian coordinate system, also referred to as a Projected Coordinate System (PCS) in some GIS software, can represent many possible locations at many possible scales. The two most common representations that are widely used in North America are the State Plane Coordinate System (SPCS) and the Universal Transverse Mercator Coordinate System (UTM) [4].

State Plane: The state plane coordinate system is a set of 126 geographic zones that cover the United States of America. Each zone is designed specifically for a particular region which allows for simple calculations and is reasonably accurate within each zone. In the SPCS coordinates are always positive inside each zone. The SPCS can be based off of NAD 27 or NAD 83 datums, and the coordinates are represented and measured in feet [4].

Each state may have multiple state plane zones with each zone strategically placed to minimize the amount of error within each zone. Additionally, each zone provides a common coordinate reference for horizontal coordinates. Depending on the shape of the state, the state plane zone can be based on two types of map projections, the Lambert Conformal Conic, or the Transverse Mercator (see Section 9.3 for more information on projections) [4].

Figure 9.6 shows the state plane zones in the 48 lower contiguous states of the United States of America. Notice that most states have multiple zones and that zones typically either run north to south, or east to west. Also note that most of the smaller states, particularly in the New England area, only have one state plane zone covering the entire state, while larger states such as California and Texas, have multiple zones so that the error can be minimized throughout the state. There are notable exceptions to this such as Montana, Nebraska, and Tennessee. State plane zones are typically designated with the name of the state followed by a section indicator, such as North, Central, and South, or a combination of those, or, West, Central, and East. An exception to this is California which specifies each zone using a different number [4].

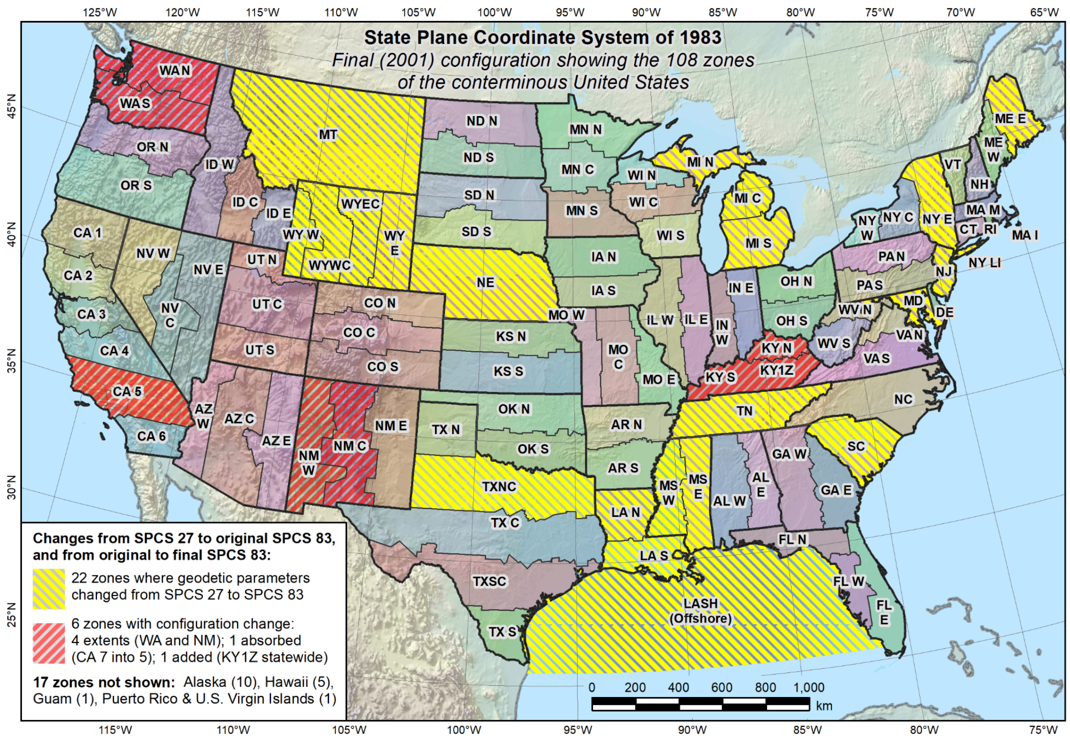


Figure 9.6. State Plane Coordinate System Zones.
Credit: Final SPCS 83 (as of 2001), *National Geodetic Survey*, public domain

Universal Transverse Mercator (UTM) Coordinate System: The UTM Coordinate System (UTMCS) is a worldwide 2-D coordinate system that splits the world into 60 zones. This system is useful because it provides for simple calculations and manages errors within each zone. Unlike the SPCS which is measured in feet, the UTM coordinate system is specified and measured in meters (cartography). In the UTM coordinate system, the Earth's surface is divided in rectangular regions, and for each of them, a different projection and a different set of geodetical parameters are used. It uses a single ellipsoid, WGS-84 [5], and a single projection, the Transerve Mercator (see Section 9.3 for more information on projections).

The UTM grid contains 60 zones, each covering 6° of longitude in width at the equator. Each zone is also segmented into 20 latitude bands, ranging from 80° South to 84° North (see Jaworski http://www.jaworski.ca/utmzones.htm [opens in new tab] for a map of the UTM zones). To avoid negative coordinate numbers, the origin is assumed to have an X coordinate of 500,000 meters and a Y coordinate of 10,000,000 meters, causing all coordinates referred to it to have only positive values [5] which are referred to as false eastings and false northings, respectively. The remaining 6° north and 10° north are covered by a different coordinate system, the Universal Polar Stereographic (UPS) coordinate system. See Figure 9.7 for a map of the UTM Zones of the contiguous United States.

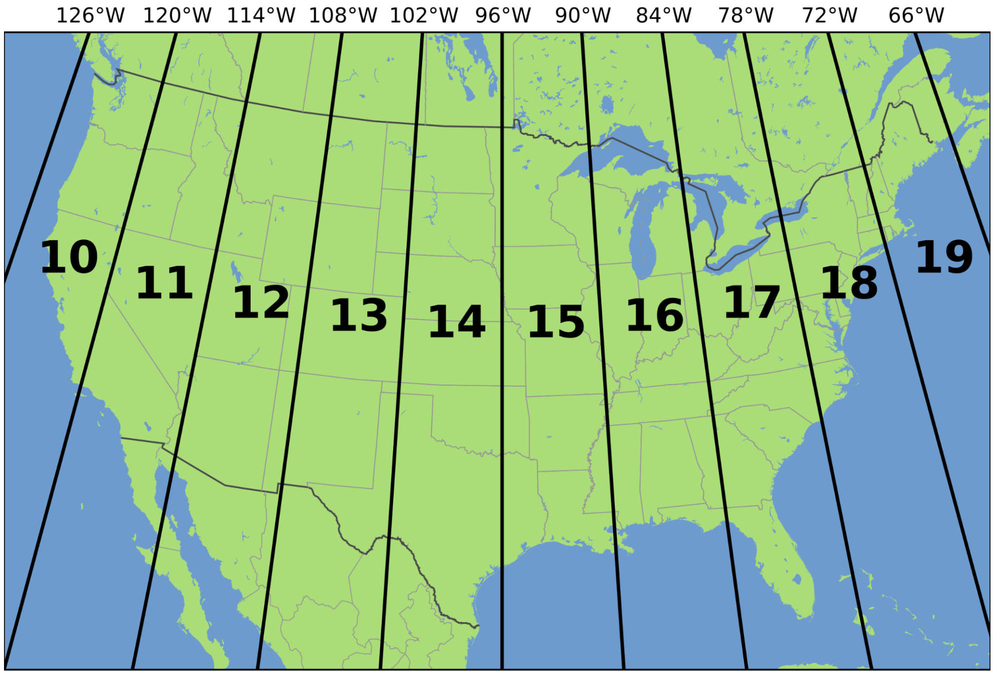


Figure 9.7. UTM Coordinate System Zones for the contiguous United States.
Credit: UTM zones USA, *Wikipedia*, Chrismurf, CC BY 3.0

9.3 Projection Concepts

Defined, a map projection is a systematic rendering of locations from the curved Earth surface onto a flat map surface. This allows us to flatten the curved surface onto a flat surface such as a piece of paper, or a computer screen. A map projection is broadly composed of three parts, the ellipsoid (see section 9.0), which models the shape of the Earth, a light source which is used to project features on the Earth's surface, and a developable surface commonly a flat piece of paper onto which the Earth's features are projected, and flattened to be used as a map [4].

To illustrate the concept of a map projection, imagine that we place a light bulb in the center of a translucent globe (Figure 9.8). On the globe are outlines of the continents and the lines of longitude and latitude called the graticule. When we turn the light bulb on, the outline of the continents and the graticule will be "projected" as shadows on the wall, ceiling, or any other nearby surface. This is what is meant by map "projection" [1].

Figure 9.8. Map Projection Conceptualization.
Credit: The Concept of Map "Projection", *Essentials of Geographic Information Systems*, Campbell and Shin,

Developable Surfaces: A developable surface is a geometric surface on which the curved surface of the Earth is projected; the end result being what we know as a map. Geometric forms that are commonly used as developable surfaces are planes, cylinders, and cones (Figure 9.9) [4]. Naming conventions for many map projections include the developable surface as well as its orientation. For example, as the name suggests, "planar" projections use the plane, "cylindrical" projections use cylinders, and "conic" projections use the cone [1].

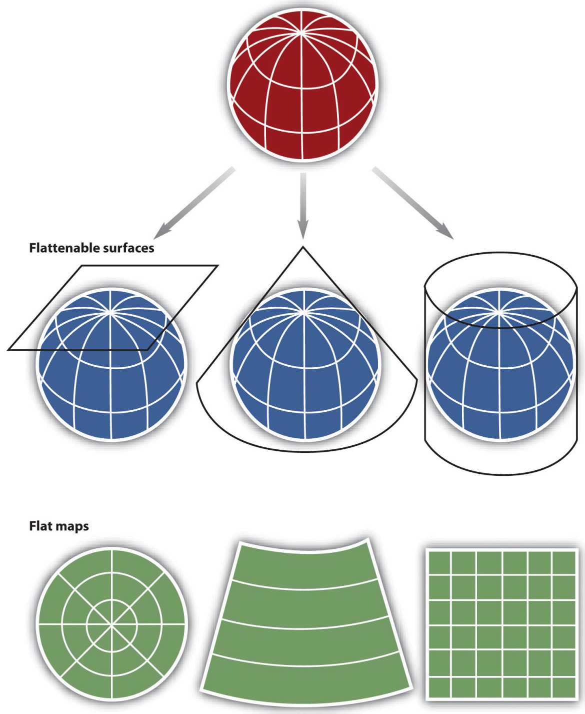


Figure 9.9. Map Projection Surfaces

Credit: Map Projection Surfaces, *Essentials of Geographic Information Systems*, Campbell and Shin,

For cylindrical projections, the "normal" or "standard" aspect refers to when the cylinder is tangential to the equator (i.e., the axis of the cylinder is oriented north–south). When the axis of the cylinder is perfectly oriented east–west, the aspect is called "transverse," and all other orientations are referred to as "oblique." Regardless of the orientation or the surface on which a projection is based, a number of distortions will be introduced that will influence the choice of map projection [1].

Tangency: The developable surfaces will interact in a few different ways with the ellipsoid. Generally, the developable surfaces will touch the ellipsoid in either two places which creates two secant lines or at a single location which creates a single tangent line or point [4]. In Figure 9.10 the tangent line of this conic projection is touching one of the mid-latitudes while the secant case is touching two lines of latitude. Figure 9.11 illustrates the tangency and secant case of a cylindrical projection, while Figure 9.12 shows the typical point of tangency for a planar projection. Understanding where the line(s) or point of tangency is important as this is the location on the map with the least amount of distortion. As you move away from the point or line(s) of tangecy, distortion increases.

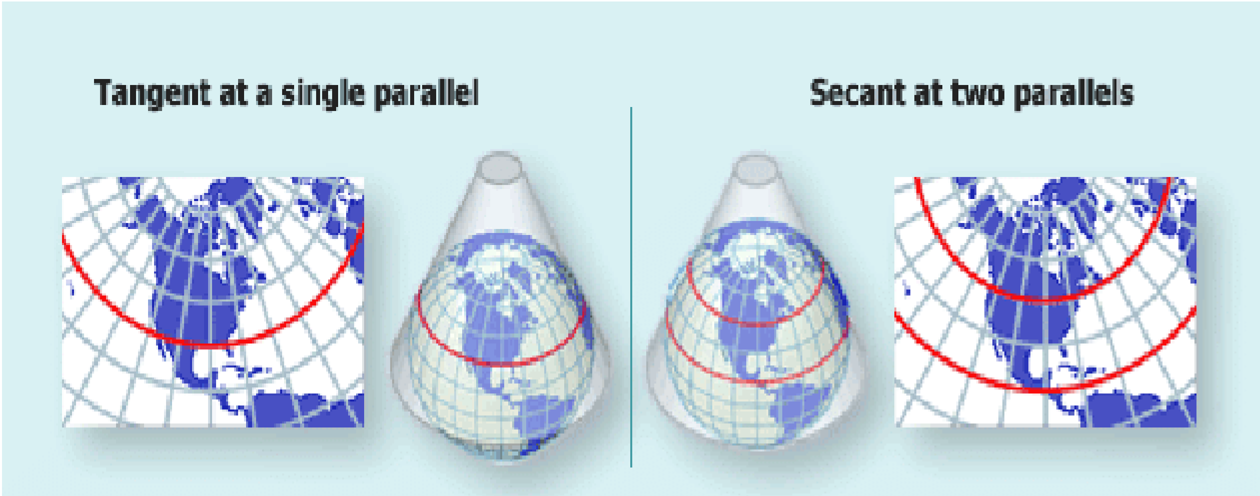


Figure 9.10. Tangent case (left) and secant case (right) of tangency for a conic projection.
Credit: Images by *USGS National Atlas*, public domain

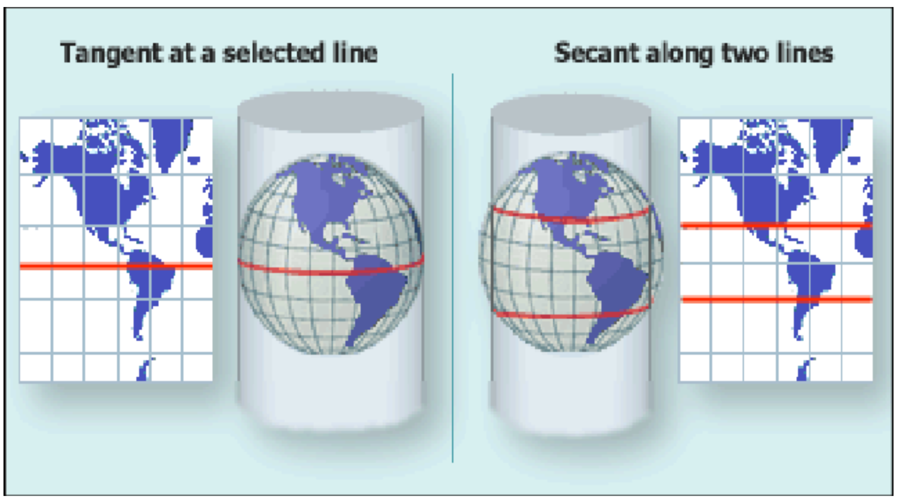


Figure 9.11. Tangent case (left) and secant case (right) of tangency for a cylindrical projection.
Credit: Images by *USGS National Atlas*, public domain

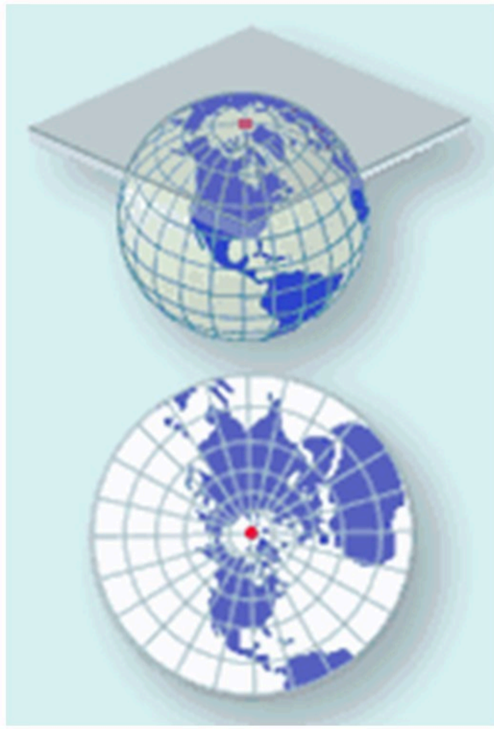


Figure 9.12. Point of tangency for a planar projection.
Credit: Image by *USGS National Atlas*, public domain

Map Projection Parameters: There are five map projection parameters, standard points and lines, projection aspect, central Meridian, latitude of origin, and light source location. Together these five map projection parameters allow an individual to selectively display and distort the Earth to create a map that is suitable to the map purpose [4].

Standard Points and Lines: Defined, a standard point or line is a point or line of intersection between the developable surface and the spheroid or ellipsoid. In the case of a secant intersection, there will be two standard lines that would define where the developable surface intersects with the spheroid. If the developable surface intersects the spheroid along a line of latitude it is known as a standard parallel. Additionally, if a standard line falls on a line of longitude it is known as a standard Meridian. When defining a map projection you must define the standard points and lines. It is not uncommon to have a map projection follow a standard parallel or standard Meridian [4].

Secant intersection between the developable surface and the spheroid can help minimize distortion over a large area by providing more control than a tangent intersection at a single point or line. Therefore placement of standard points and lines (tangency) is one of the most important parameters to consider when defining a map projection [4].

Projection Aspect: Aspect refers to where the developable surface is placed relative to the globe. Each developable surface has what is considered a normal, transverse, and oblique aspect (see projection aspects by developable surface - https://eipd.dcs.wisc.edu/for-credit/GIS-cert/images/geog370/L2_Images_Projections/Geog370_L2_F24.jpg [opens in new tab]).

Light Source Location: There are three hypothetical light source locations for planar projections in reference to the globe - gnomonic, stereographic, and orthographic (Figure 9.13). In the gnomonic light source position, the light source is placed at the center, or core, of the Earth. The light is then projected through the Earth's surface and projects the landmasses onto the developable surface [4].

In the stereographic projection, the light sources are placed at the opposite side of the Earth from where the developable surface has its secant or tangent intersection. In this case, we see a less severe differential between where the Earth is compressed and elongated, but no location is free from distortion [4].

In the orthographic position, the light is placed at a theoretically infinite distance from Earth opposite from the point of intersection or tangency. This allows formidable distortion in the center of the projection; however, there is significant compression at the extremities of the map [4]. Orthographic is how the Earth is seen from space so is often used to portray the globe as such.

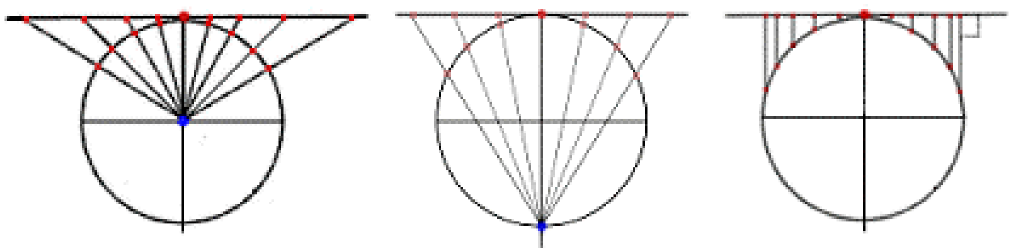


Figure 9.13. The different light sources for planar projections, gnomonic (left), stereographic (middle), orthographic (right).

Credit: Image by Lantmäteriet, *Land Survey, Ministry of Rural Affairs and Infrastructure, Sweden*, public domain

Great Circle: A great circle is a circle that results when a plane intersects the Earth going through the center effectively dividing the Earth into two equal parts. Based on this definition, all meridians and the equator are considered to be great circles as they all divide the Earth into two equal halves. An arc segment of a great circle is the shortest path between two points on a spherical surface (Figure 9.14). Therefore, to determine the shortest distance between two points on Earth you would first construct a great circle and then calculate the distance along that segment [4].

Rhumb Line: A Rhumb Line is a curve that crosses each Meridian at the same angle (Figure 9.14). This is also known as a loxodrome. Rhumb lines are useful for navigation because they follow a constant bearing or azimuth. While the Rhumb line is useful for navigation, it does require that you follow a longer distance rather than the shortest possible distance, a great circle [4].

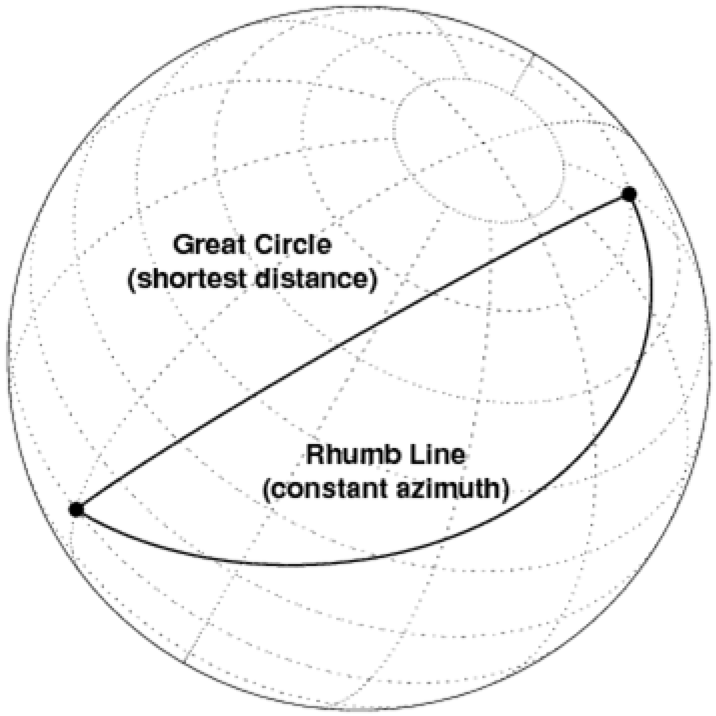


Figure 9.14. Great circle (shorter distance) and rhumb line (constant compass bearing).
Credit: Rhumb line, Introduction to Cartography, Ulrike Ingram, CC BY 4.0

9.4: Map Projections

Cartographers have developed hundreds of map projections, which are grouped into three main families based on the type of geometric shape that is used to transform a sphere into a plane. These families are called developable surfaces and consist of cylinders, cones, and planes. A number of other map projections are based on a polyhedron (e.g., Fuller-Dymaxion), though they are not described here.

Planar Family: The planar or azimuthal map projection family is when a spheroid is projected onto a flat plane. Based on this interaction between the spheroid and a developable surface we see distortions increasing in a series of concentric bands from the point of tangency. The planar family of map projections is commonly used to display larger-scale maps and polar regions of the Earth. The planar family's normal aspect is polar [4].

Cylindrical Family: The cylindrical family projects the spheroid onto a cylinder that when rolled flat is a rectangle. The spheroid is distorted with increasing exaggeration towards the outer edges of the map, typically the north and south poles, as the standard aspect for a cylindrical map is equatorial. The cylindrical map projection family is commonly used to display the entire world, though can beused for maps at any scale. One unique characteristic of the cylindrical family is that all parallels and meridians intersect at 90° angles [4].

Conic Family: The conic family uses a cone onto which the spheroid is projected. Distortion is present in concentric bands moving outward from the line of tangency which the standard is a parallel in the mid-latitudes and the normal aspect is polar. The conic family is commonly used to display areas of Earth having a greater east-west extent [4].

9.5 Map Projection Distortions

All projected maps alter the area, shape, distance, and/or direction of a map or its features. Map projection properties exist because the conversion of a three-dimensional object, such as the Earth, to a two-dimensional representation, such as a flat paper map, requires the distortion of the three-dimensional object. The three-dimensional spherical surface is torn, sheared, or compressed to fit the developable surface. The four map projection properties described can either be held true or distorted [4].

Area and shape are considered major properties and are mutually exclusive. That means, that if area is held to its true form on a map, shape must be distorted, and vice versa. Distance and direction, on the other hand, are minor properties and can coexist with any of the other projection properties. However, distance and direction cannot be true everywhere on a map [4].

Compromise projections do not hold a single map projection property true. Their goal is to simultaneously minimize all map distortion properties, but may not hold any of the four map projection properties as true.

9.6 Projection Examples

The map projections discussed below are done so by the property in which the projection holds true (does not distort).

Equal Area Map Projections: The equal area map projection, also known as the equivalent map projection, preserves the area relationships of all parts of the globe. You can easily identify most equal area map projections by noting that the meridians and parallels are not at right angles to each other. Additionally, distance distortion is often present on equal area map projections, and the shape is often skewed [4].

Equal area maps are useful for quantitative thematic maps when it is desirable to retain area properties, such as for dot density maps. This is especially useful for choropleth maps when the attribute is normalized by area. Holding area true allows comparison of density between different enumeration units such as counties. One example of an equal area projection is the Lambert Cylindrical Equal Area projection (Figure 9.15) [4].

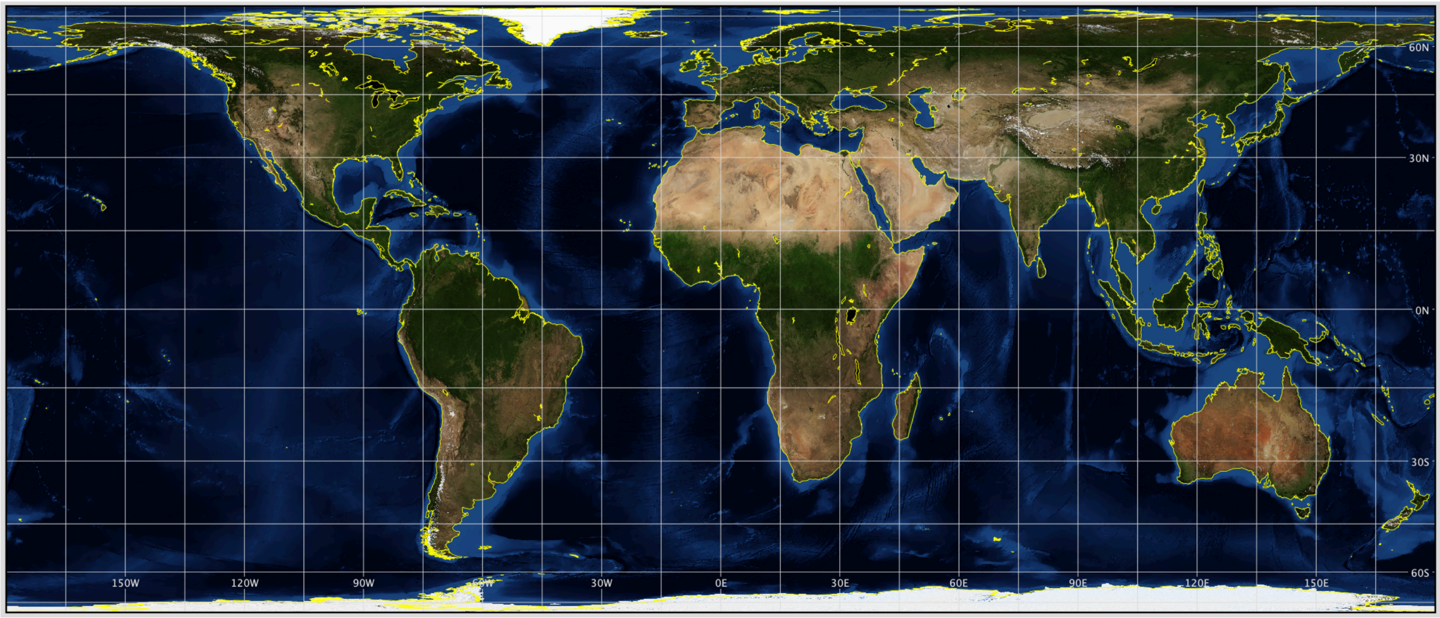


Figure 9.15. Lambert cylindrical equal-area projection
Credit*:* Image from Map Projection Explorer, NASA, public domain

Conformal Map Projections: Conformal map projections preserve angles and thus the shape of the map features. These projections can usually be identified by the fact that meridians intersect parallels at right angles and areas are distorted significantly [4]. Conformal projections should be used if the main purpose of the map involves measuring angles or representing the shapes of features. They are very useful for navigation, topography (elevation), and weather maps [6].

The Mercator projection (Figure 9.16), perhaps the most famous of all map projections, is a conformal map projection that preserves shape. However, notice the massive amount of distortion in the lower latitudes towards the South Pole and the northern latitudes near the North Pole. Also, note that the parallels and meridians intersect at 90° angles [4].

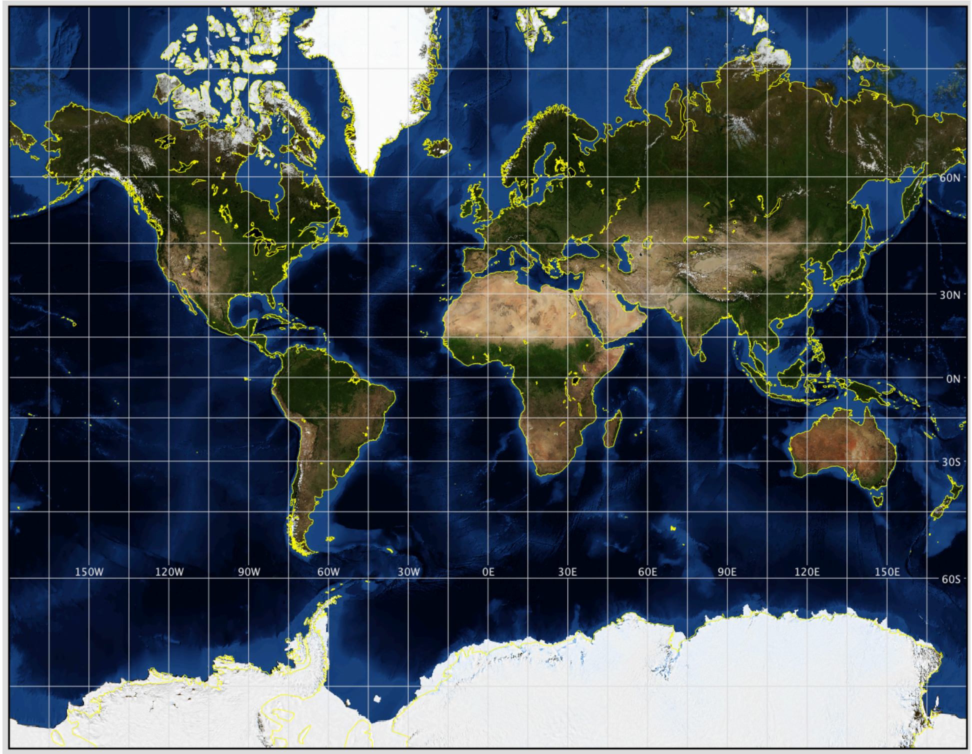


Figure 9.16. Mercator Map Projection
Credit*:* Image from Map Projection Explorer, [NASA](), public domain

Equidistant Map Projections: The equidistant map projection aims to preserve great circle distances which means a distance can be held true from one point to all other points or from a few select points to others but not from all points to all other points. Identifying marks of the equidistant map projection are that they are neither conformal nor equal area, and tend to look less distorted (Figure 9.17). Equidistant map projections are useful for general-purpose maps and Atlas maps [4].

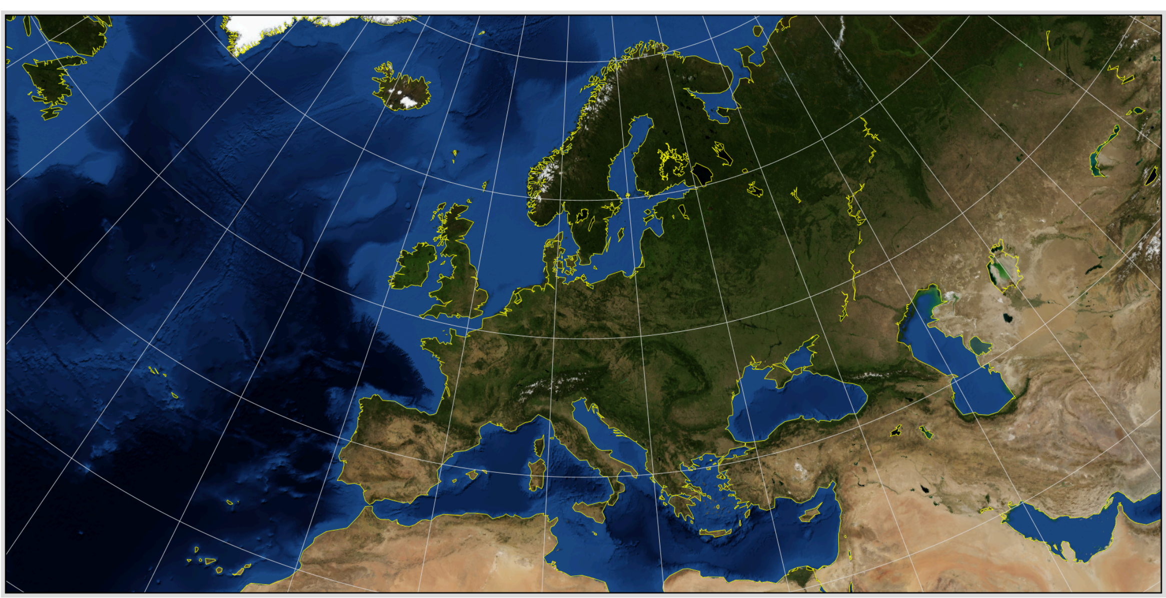


Figure 9.17. Equidistant Conic Projection
Credit*:* Image from Map Projection Explorer, [NASA](), public domain

Azimuthal Map Projection (Equidirectional): The azimuthal map projection (Figure 9.18), also known as the true direction map projection, preserves direction from one point to all other points in the map (shape, area, and distance are distorted). It is important to note that direction is not true between non-central points. Direction is only true when measured to or from the specific points chosen to have true direction. Azimuthal map projection is often used for navigation [4].

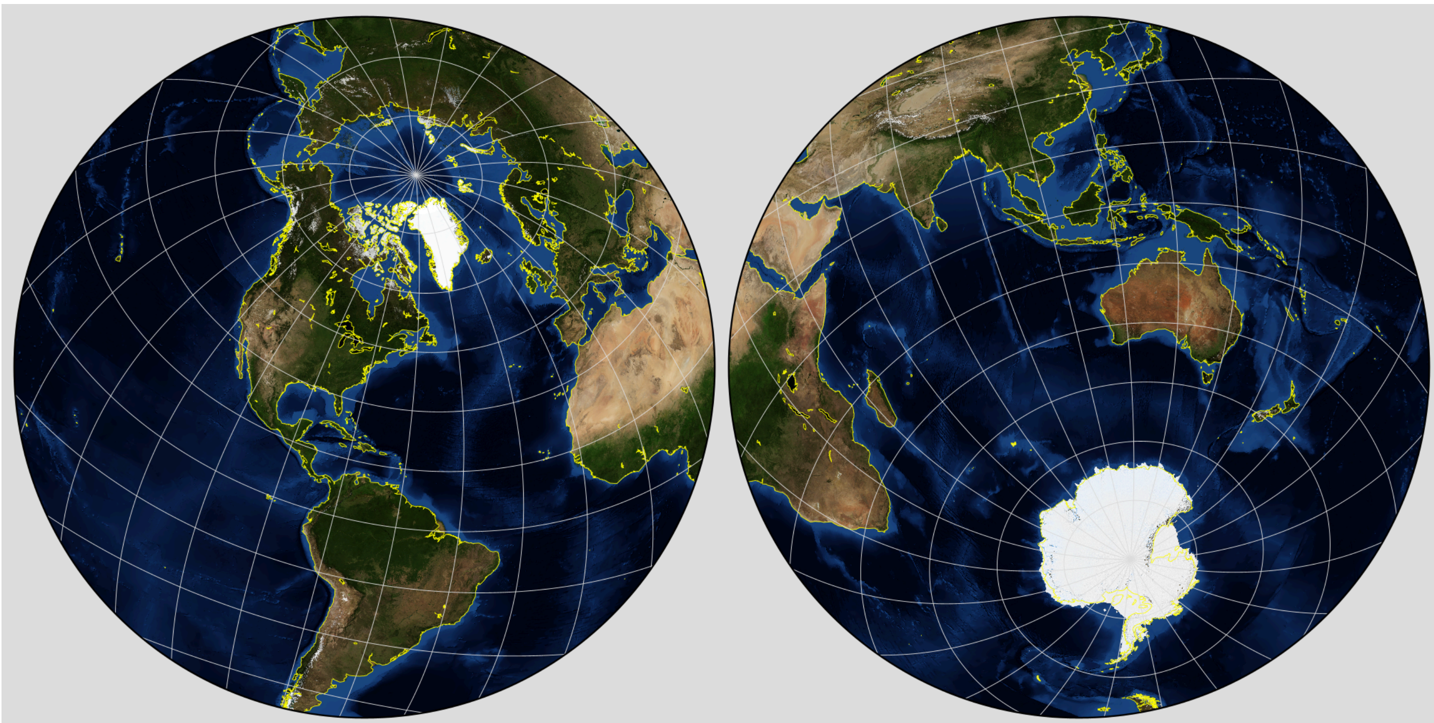


Figure 9.18. Azimuthal Equidistant Map Projection
Credit*:* Image from Map Projection Explorer, NASA, public domain

Compromise Map Projection: A compromise projection preserves no one property but instead seeks a compromise that minimizes distortion of all kinds, as with the Robinson projection (Figure 9.20), which is often used for small-scale thematic maps of the world [6]. Compromise projections, in general, do a reasonable job of showing the true shape, distance, direction, and area of the features of the Earth.

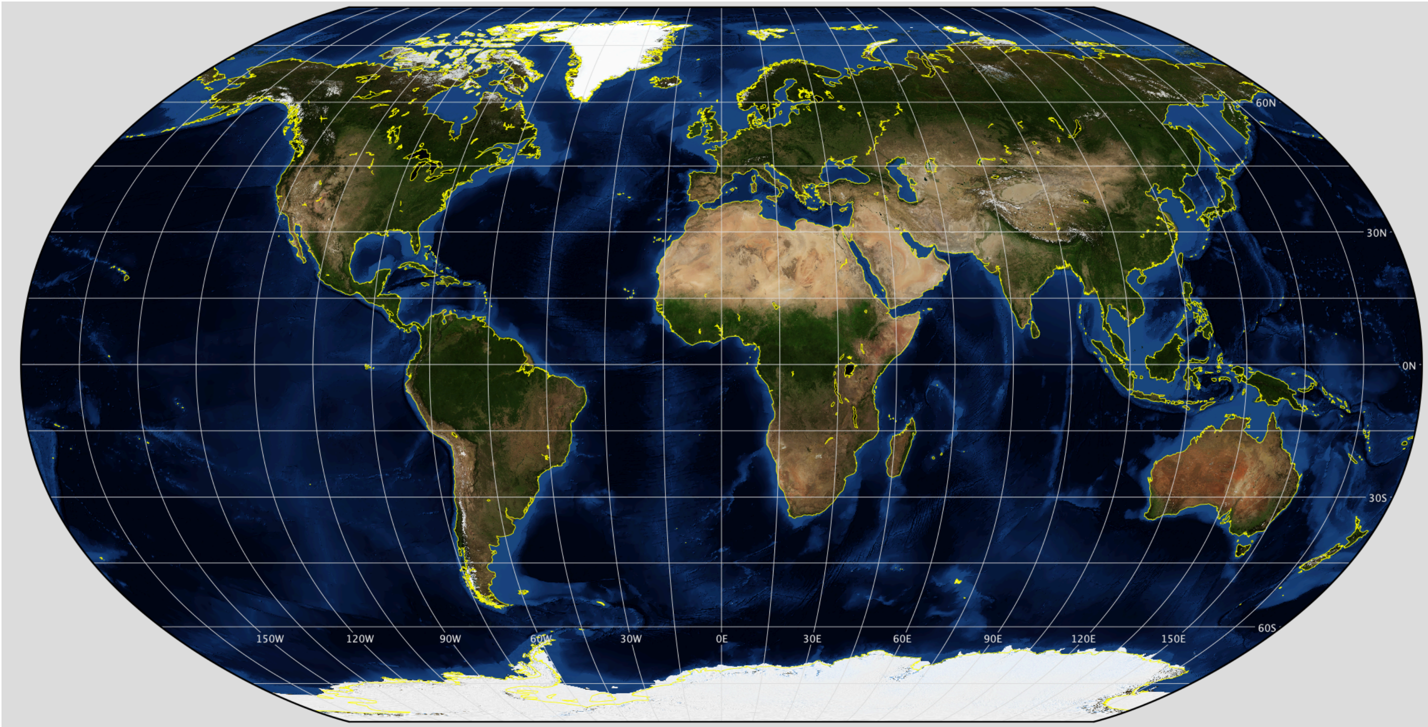


Figure 9.20. Robinson Projection
Credit*:* Image from Map Projection Explorer, NASA, public domain

9.7 Determining Distortions

A common method used to determine distortions on a map projection is to apply Tissot's indicatrix. It is a concept developed by French cartographer Nicolas Auguste Tissot [7]. All maps distort at least one of the projection properties of shape, area, direction, and distance. Tissot's indicatrix helps to quantify the distortion of the projection properties of shape and area as shown on the map.

Tissot's indicatrix is composed of small circles centered at points on the Earth, typically at the intersection of lines of latitude and longitude. The Earth is then projected onto a map using a projection. We then consider the shape and area of the circles across the map surface to determine the amount and distribution of distortion [4].

There are three ways in which we can visualize distortions using Tissot's indicatrix - size/area change, shape change, or both (Figure 9.21).

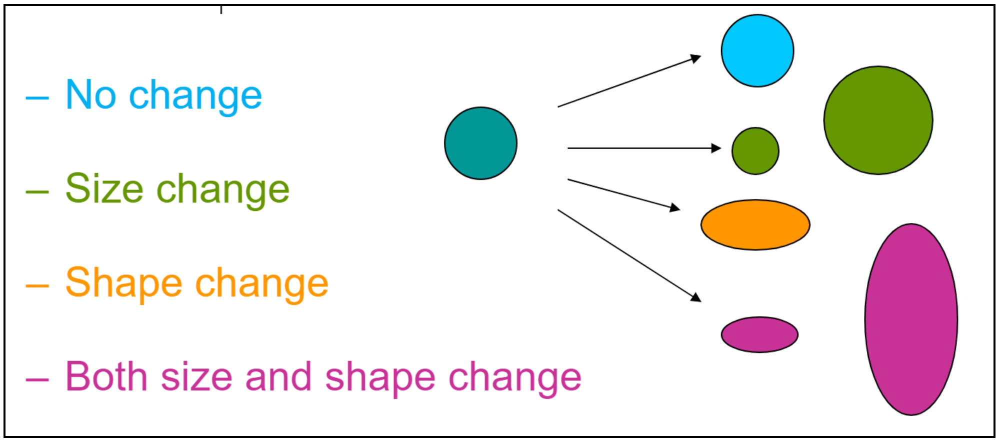


Figure 9.21. Changes that can be visualized using Tissot's indicatrix.
Credit: Image created by Barbara Buttenfield, University of Colorado Boulder, used with permission

Conformal Projection: On a conformal map projection the circles will continue to be perfect circles, not changing in shape, but their size/area will vary over the map such as with the Mercator Projection (Figure 9.22) [4].

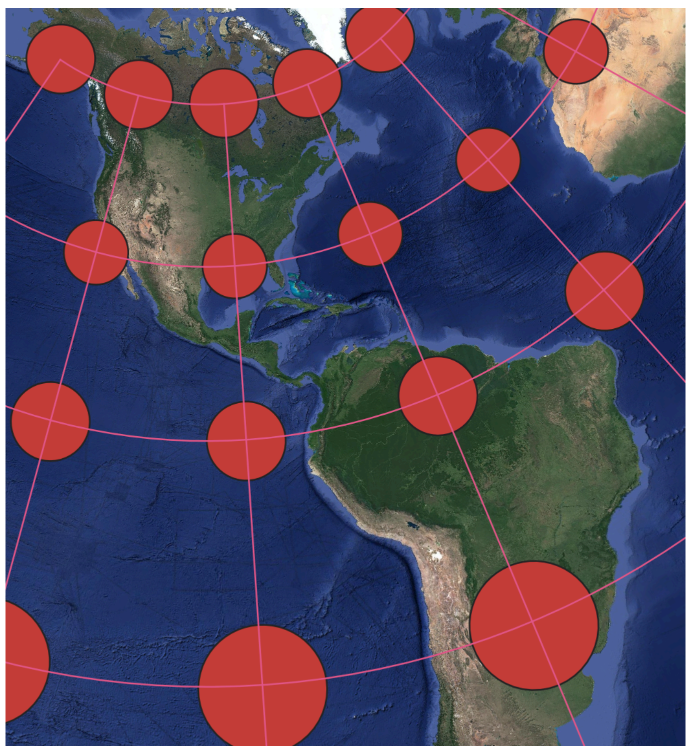


Figure 9.22. Lambert conformal conic projection with Tissot's indicatrix showing size/area change only.
Data Source: Google Satellite Imagery accessed through QGIS.

Equal-Area Projection: On an equal area map projection the circles will be transformed into ellipses but the area of the ellipses will be the same as the area of the original circles. That means although they will change shape they will not change in size/area such as with the World Equal Area projection (Figure 9.23) [4].

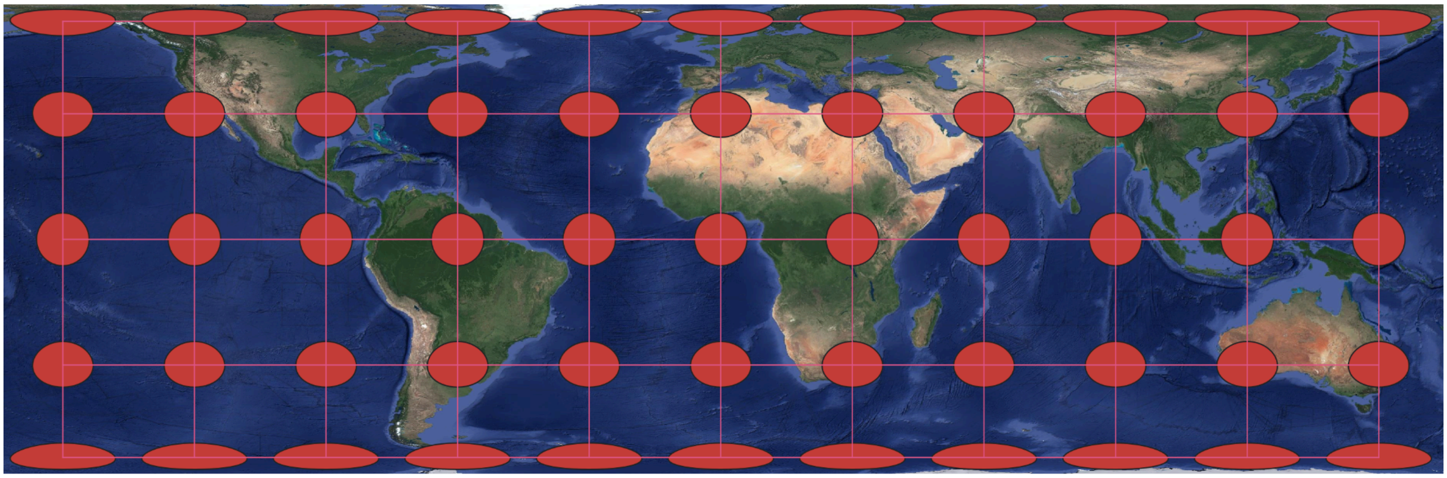


Figure 9.23. World Equal Area cylindrical projection with Tissot's indicatrix showing shape change only.
Data Source: Google Satellite Imagery accessed through QGIS.

Compromise Projection: For a compromise projection, no property is preserved as it seeks to minimize distortion of all the map properties. Therefore, as can be seen on the Robinson projection (Figure 9.24), neither size/area nor shape is retained as evidenced by Tissot's indicatrix [4].

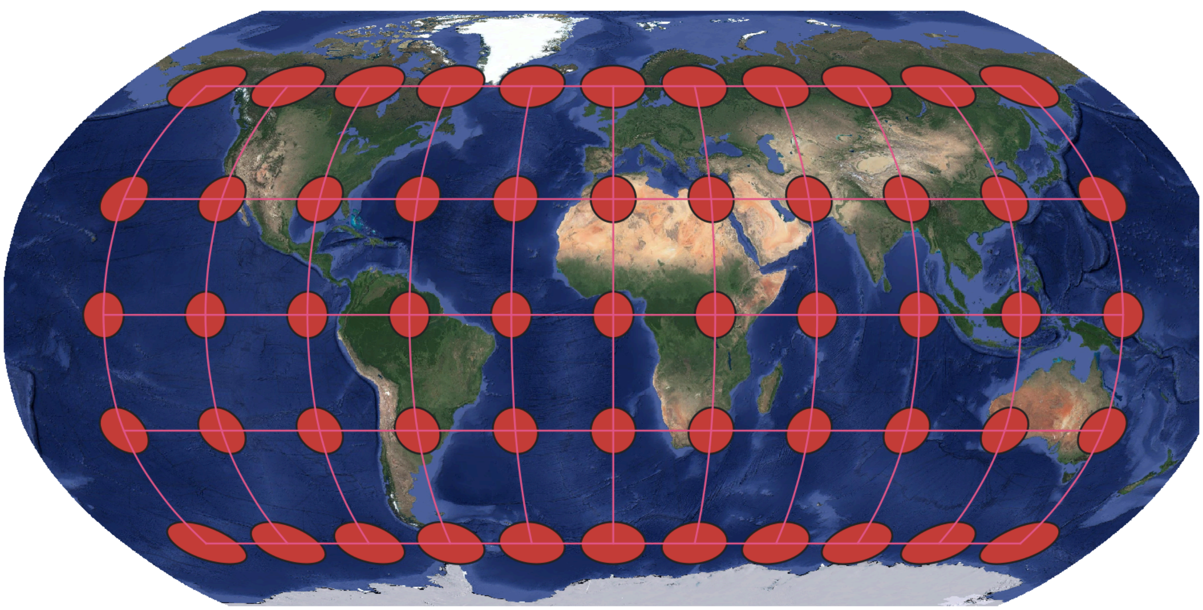


Figure 9.24. Robinson projection with Tissot's indicatrix showing both shape and size/area change.
Data Source: Google Satellite Imagery accessed through QGIS.

Equidistant and Equidirectional Projections: Both equidistant and equidirectional map projections sacrifice shape and area representations to preserve either distance or directional measurements, respectively. This can be observed on an equidistant conic projection (Figure 9.25) as neither size/area nor shape is retained as evidenced by Tissot's indicatrix [4].

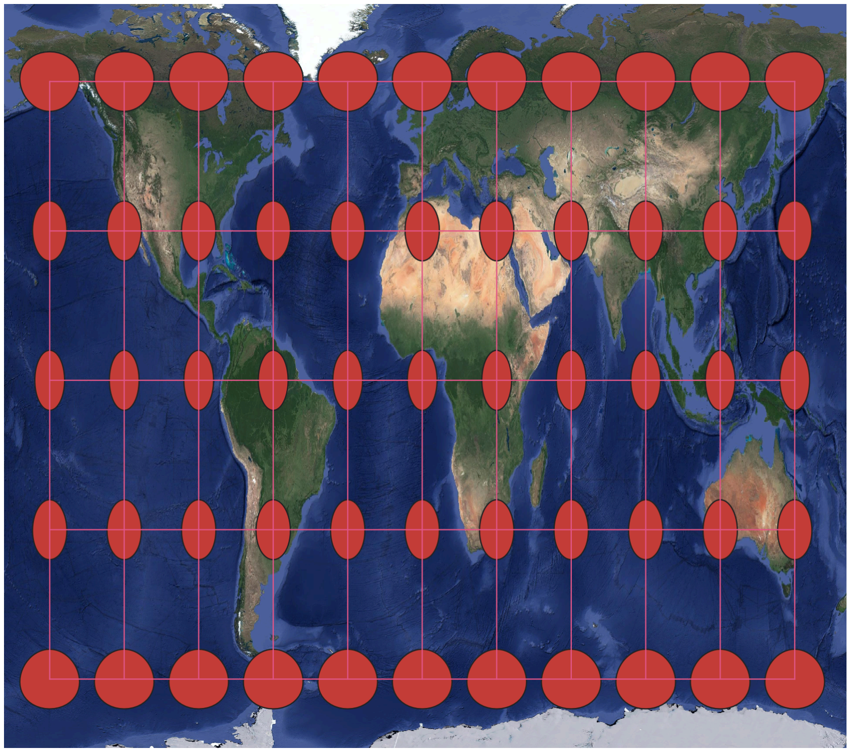


Figure 9.25. Equidistant projection with Tissot's indicatrix showing both shape and size/area change.
Data Source: Google Satellite Imagery accessed through QGIS.

References - materials are adapted from the following sources:

[1] Essentials of Geographic Information Systems by Saylor Academy under a CC BY-NC-SA 3.0 license

[2] GEOG 486 Cartography and Visualization by Cary Anderson, Pennsylvania State University, under a CC BY-NC-SA 4.0 license.

[3] Geographic Information Systems and Cartography by Adam Dastrup under a CC BY-NC-SA 4.0 license

[4] Introduction to Cartography by Ulrike Ingram under a CC BY 4.0 license

[5] Introduction to GIS by Victor Olaya under a CC BY 3.0 license

[6] Mapping, Society, and Technology by Steven M. Manson under a CC BY-NC 4.0 license

[7] Wikidata under a CC0 license

Chapter 10
Data Classification

10.0 Choropleth Maps

Choropleth maps show relative magnitude or density per a specified enumeration unit such as a county or state. Alternatively, isoline maps, also known as isarithmic maps, use lines to connect point locations with similar values, instead of using enumeration units. When portraying information through choropleth maps, it is typically generalized by the grouping of the data into different classes and different numbers of classes, a process referred to as classification [1].

Normalization of Data: Raw and total values are often useful on maps, however, areas that are larger may have more of some value than areas that are smaller in size simply because there is more area. For example, based on purely its size California can hold more people than Rhode Island. In other words, if you want to compare the value of a variable between two areas that vary in size, a fair comparison is not possible. Therefore, we normalize data to allow for meaningful comparisons of values [3].

Raw vs. Normalized Data: If we compare raw data versus normalized data by looking at the total population of the United States we get two completely different-looking maps. Figure 10.0 displays the total population by state as a choropleth map. The lighter colors are where there is less population and the darker colors are where there is more population. Note that California and Texas are the states with the most pepople and are also the largest states so they naturally can have higher populations. However, if we normalize the data by area we can get a view of how dense each state is with respect to population. Figure 10.1 shows the population per square mile which shows a significantly different view than the total population map. Here, the smaller states in the Northeast have less total population but more people per square mile because the states are significantly smaller than the states out West [3]. By portraying the values as densities, the data has been normalized, thereby accounting for the size variations present in the enumeration units.

Figure 10.0: Classified Choropleth Map of Total Population by State (raw data)
Credit: Choropleth Map of Total Population by State, *Introduction to Cartography*, Ulrike Ingram,

Figure 10.1: Classified Choropleth Map of Population Density (normalized data)
Credit: Choropleth Map of Population Density per Square Mile, *Introduction to Cartography*, Ulrike Ingram,

Ratios, Proportions, and Rates: A rate is the number of items in one entity divided by the number in a second entity, such as population density as we saw in the example above. Proportions represent the relationship of a part of a whole and are represented as percentages. For example, the population of Colorado is 49% female. Rates express the relationship as a value per some much larger value (death rate = 10 per 1,000 people). All of these are forms of data normalization that can be used when mapping aggregated data [1].

10.1 Descriptive Statistics

Descriptive statistics are ways to describe data sets quantitatively, using measures of central tendency and dispersion. Cartographers use descriptive statistics to explore the characteristics of the data they are mapping.

Central Tendency: Central tendency describes the distribution of the data and focuses on summarizing the data using one particular value representing the "center" of the data. Three measures are commonly used to describe the central tendency: mean, median, and mode.

The **mean** is the average value of the data set. The average value is defined by summing all values of the data set and dividing it by the number of values in the data set. The **median** represents the middle point of the data. For instance, if we had three observations in our data set we would sort the observations in ascending order and then look at the value at the midpoint which would be the second observation in this case. If there is an even number of observations in a data set, then the median will be the average of the two most central observations. The **mode** is the most common value found in the data set. If multiple values are tied as the most common value, the data set has multiple modes [3]. For instance, in a data set with the values of 1, 2, 2, 3, 10, 12, 12, 15, and 16, the modes are 2 and 12 since they are tied for the most common value.

Dispersion: Dispersion measures the variability of the data set. There are two measures of dispersion, variance and standard deviation. Both measure the spread of the data around the mean, summarizing how close, or far, each observed data value is to/from the mean value.

The **variance** takes the sum of the squares of the deviations divided by the number of observations. The units of variance are identical to the original units of measure. So for instance, if the original units of the observations were in feet then the variance will report how much the observations vary on average in feet. The **standard deviation** is similar to the variance except that it takes the square root of the sum of the squares of the variance divided by the number of observations. The standard deviation measures dispersion in a standard way so that two different data sets can be compared. Standard deviation does not report the dispersion in the original units of the observations [3], but by deviations from the mean.

Normal Distribution: A normal distribution of data means that most of the values of the data set are close to the average value while relatively few examples tend to one extreme or the other, above or below the mean (Figure 10.2). The x-axis is the value in question and the y-axis is the number of observations for each value on the x-axis. The standard deviation tells us how tightly all the various observations are clustered around the mean of the data set. One standard deviation away from the mean in either direction on the horizontal axis accounts for about 68% of the observations in the data set. Two standard deviations away from the mean which are the four areas closest to the center account for about 95% of the observations in the data set. Three standard deviations account for about 99% of all the observations of the curve [3].

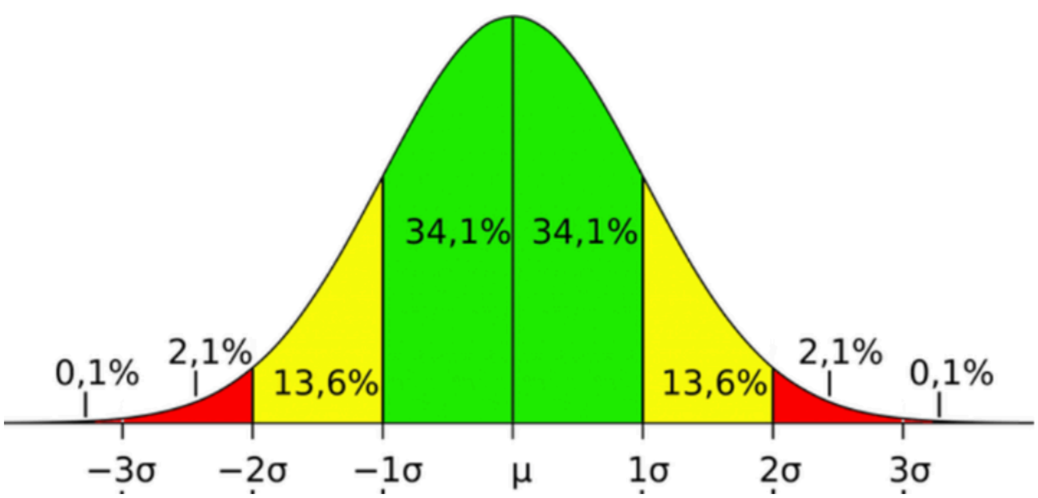


Figure 10.2. Normal distribution

Credit: Image from Characteristics of a Normal Distribution, Boston University School of Public Health, Wayne W. LaMorte, CC BY-NC 3.0

The skewness of the curve tells us whether the peak of a distribution is to one side of the mean or the other. If a data set has a negative skew then the peak of the distribution is above the average. If a data set has a positive skew the peak of the data set is below the average [3].

Outliers are data values that are at an abnormal distance from the other data values and can have a significant impact on the measures of central tendency and dispersion. However, it is important to include the data values in your map unless there is a known reason to exclude them, such as data entry inaccuracies.

10.2: Data Classification

Data classification categorizes objects based on a set of conditions into separate bins or classes. Classification may add to or modify attribute data for each geographic object. For example, a classification could reclassify an attribute of “urban” changing it to “dense”. When discussing how data is broken into classes, different classification methods are available and have a direct effect on how the map and data are perceived by the map user, therefore much care must be taken when choosing a data classification method [3].

Why Classify Data?: The primary reason to classify data is to simplify the data for visual display. For instance, if we assign a unique color to each state in the United States based on its population (referred to as an unclassed map) it is very difficult to look for patterns [3]. However, if we group the data values in some fashion (classify the data), it is easier to observe the data patterns (Figure 10.3). Classifying the data also makes the comparison between two different maps much easier. For example, mapping the population density for Colorado Counties for 1990 and 2020 using the same classification procedures will make it much easier to see where and how much the population density has changed; such comparisons would be much more difficult, if even possible, with unclassed data.

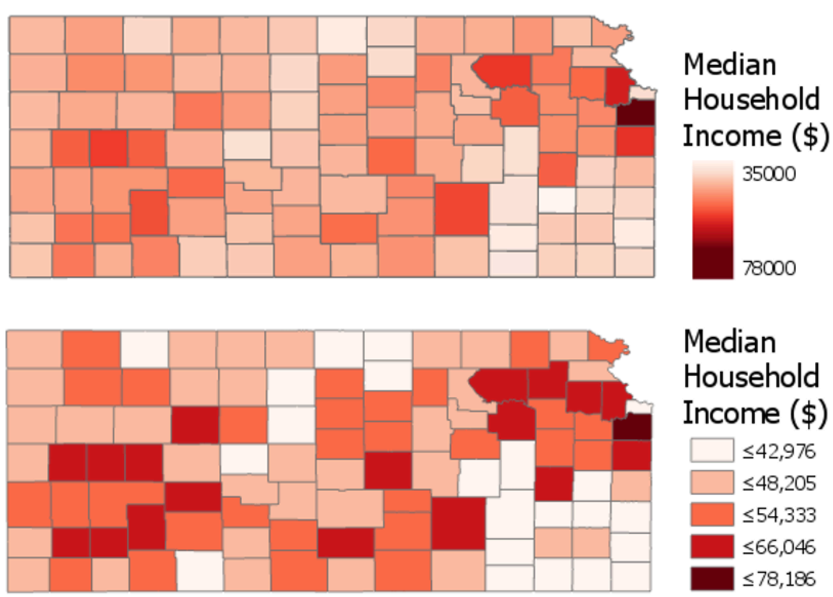


Figure 10.3. Unclassed data vs. classified data maps.
Credit: An unclassed (top) and classed (bottom) choropleth map, *Penn State University*, Cary Anderson, CC BY-NC-SA 4.0

Classification Requirements: The main consideration when classifying data is that the character of the original data is maintained even after classification. The classification method used should still represent the trend of the data. The goal in classifying data is to simplify the data for visualization, not to modify the character of the data [3]. To do this, start off by addressing the following questions [1]:

1) How many classes?
2) What classification method?
3) What pattern of error is introduced?

Choosing the Number of Classes: The number of classes that is most appropriate for a particular map is based on a variety of factors. The map audience must be considered such as their domain expertise, cognitive skills, and time available for assimilation of the map. The most important factor though, is the spatial pattern of distribution. In other words, maintaining the integrity of the dataset by showing the spatial patterns that exist. Data quality and uncertainty are factors that should not contribute to class number selection [1].

If too many classes are chosen (Figure 10.4), it requires the map reader to remember too much when viewing your map and may also make the differentiation of class colors difficult [3]. Generally speaking, distinguishing more than seven classes with the human eye becomes challenging, although using five classes is a widely accepted standard. With a well-chosen color scheme, distinctions between colors will remain clear. Conversely, more classes may reduce the statistical error that is introduced during classification [1]. The exception is for diverging color schemes since they rely on two sequences of color meeting at a central value or class. If too few classes are used (Figure 10.5), it oversimplifies the data reducing perceptual complexity but possibly hiding important patterns. Additionally, each class may group dissimilar items together [1][3].

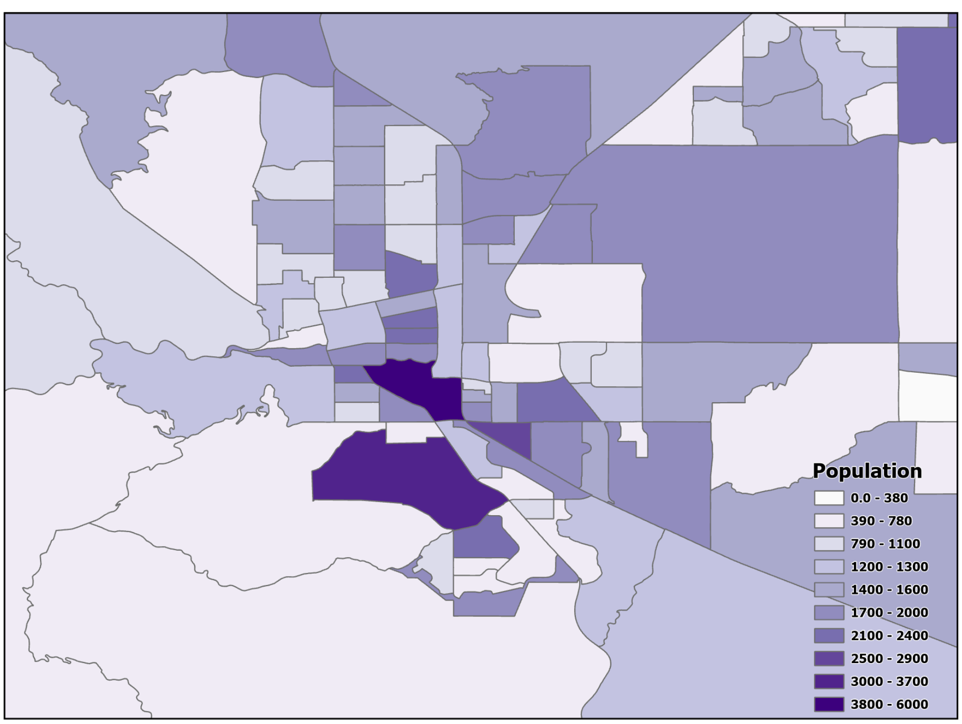


Figure 10.4. Too many classes.
Data Source: US Census Bureau

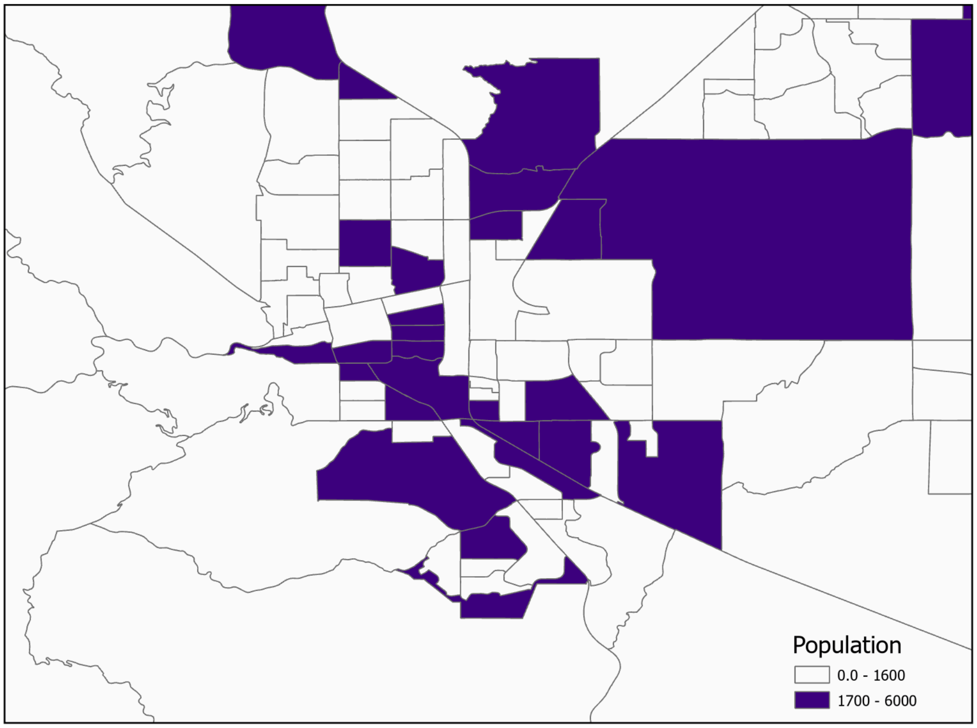


Figure 10.5. Too few classes.
Data Source: US Census Bureau

10.3. Classification Methods
A variety of classification techniques are available within GIS software. This section focuses on what those techniques are and the advantages and disadvantages of each classification method.

Binary Classification: Binary classification is when objects are placed into two classes. The two classes can be the value of 0 and 1, true and false, or any other dichotomy that you would use (e.g., Republican and Democrat). In Figure 10.6 the states were classified into two binary classes, one class for the northern states and one class for the southern states [3].

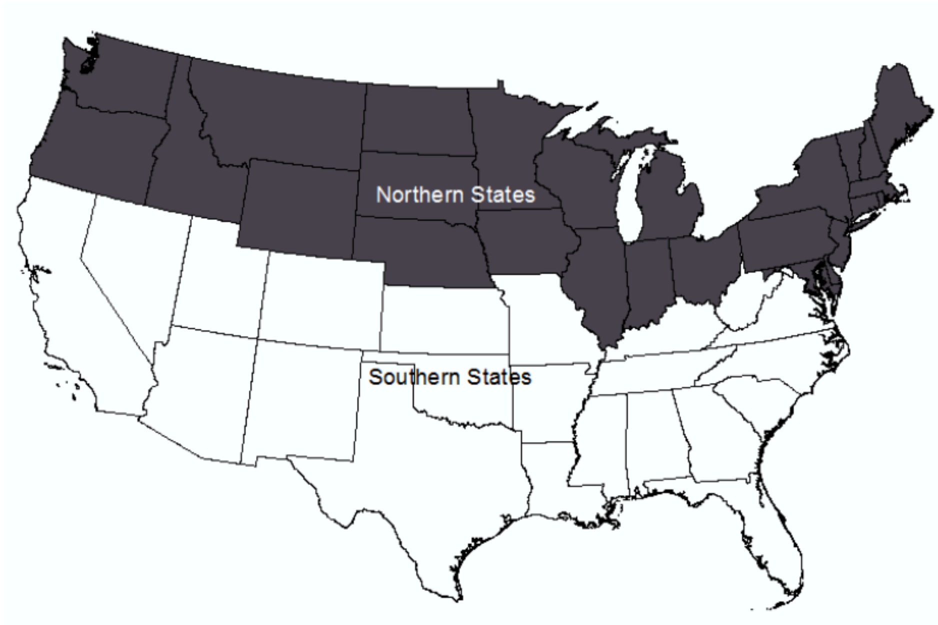


Figure 10.6. Binary classification
Credit: Two binary classes, *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

Jenks Natural Breaks Classification: The Jenks or Natural Breaks classification method (Figure 10.7) in its most simplistic form, uses breaks in a data histogram as class breaks and assumes that grouped data are alike. It uses a mathematical formula (Figure 10.8) based on the data and class variances to group like data values together (minimizing variance) while maximizing the difference between classes (maximizing variance).

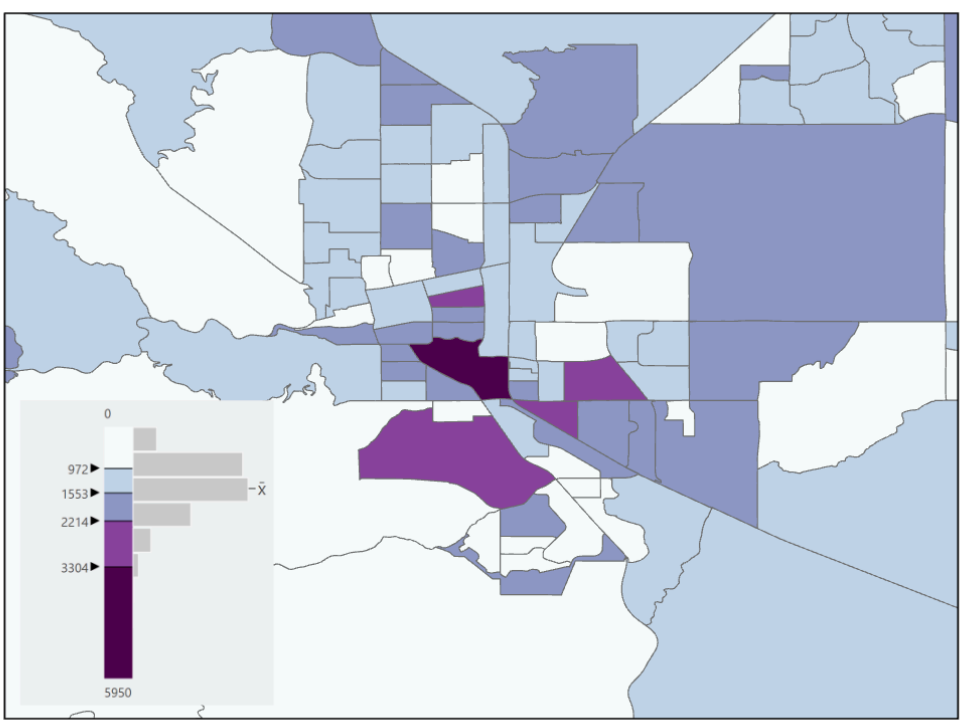


Figure 10.7. Natural Breaks classification
Data Source: US Census Bureau

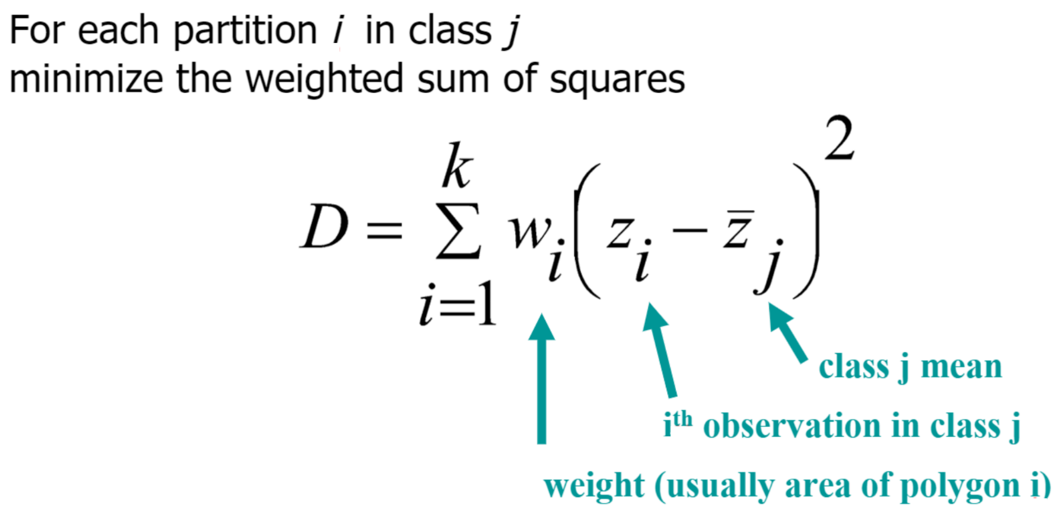


Figure 10.8. Formula used for Natural Breaks classification.
Credit: Image by Barbara Buttenfield, University of Colorado Boulder, used with permission

The advantage of the Natural Breaks classification method is that it considers the distribution of data to minimize in-class variance and maximize between-class variance [3]. The disadvantages are that it can be difficult to apply to a larger data set and may miss natural spatial clustering [1].

Standard Deviation: The standard deviation classification method creates classes relative to the mean data value by showing standard deviations above and below the mean. This classification method should only be used on normally distributed data and when the number of observations is large enough to justify using this method (3 standard deviations both above and below the mean). Since this method shows positive and negative data, a diverging color scheme should be used.

The advantage of this classification method is that it looks at the data distribution and produces constant class intervals above and below the mean. The disadvantage of this method is that most data are not normally distributed and are therefore not a good fit for this method. This method requires that the map user understands statistics to appropriately interpret the map [3].

In Figure 10.9 a diverging color scheme is used with the colors tending towards reddish-brown being observations below the mean and observations towards purple being above the mean. The data for the state has a positive skew, therefore it is not normally distributed, and the standard deviation classification method should not be used for this data set - this map is only provided as an example of the classification scheme and an appropriate color ramp for this method.

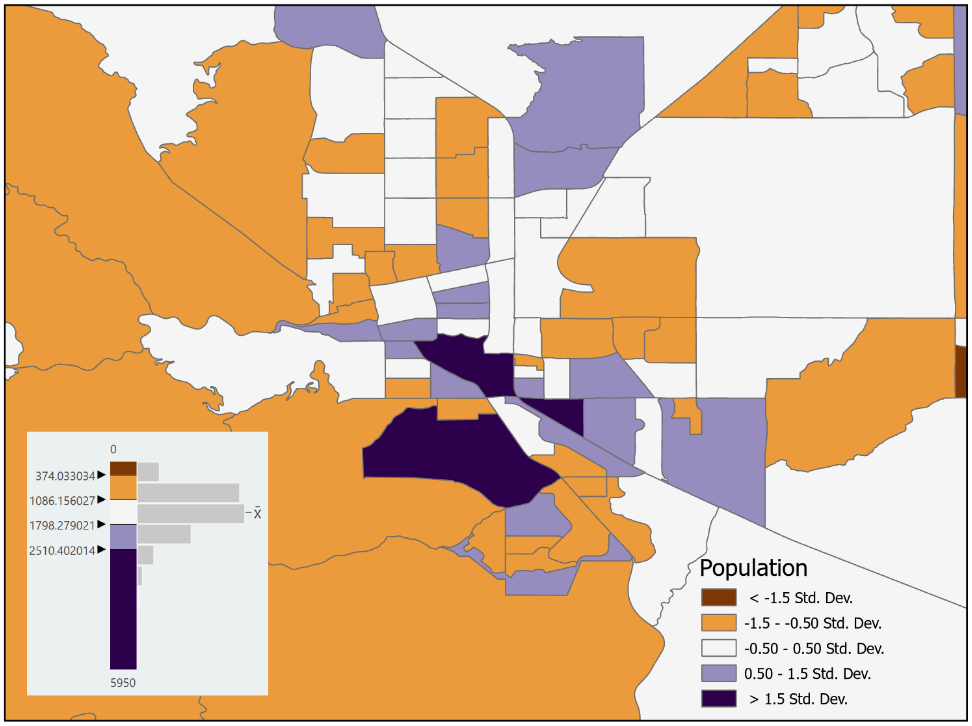


Figure 10.9. Standard deviation classification with a diverging color scheme.
Data Source: US Census Bureau

Equal Interval: The equal interval classification method (Figure 10.10) creates classes with equal ranges. The class range is calculated by taking the maximum value of the data set, subtracting the minimum value from it, and then dividing that by the number of observations in the data set. The advantages of this method are that it is easy to understand, simple to compute, and leaves no gaps in the legend. Disadvantages are that it does not consider the distribution of data and it may produce classes with zero observations (classes not used on the map, but shown in the legend) [3].

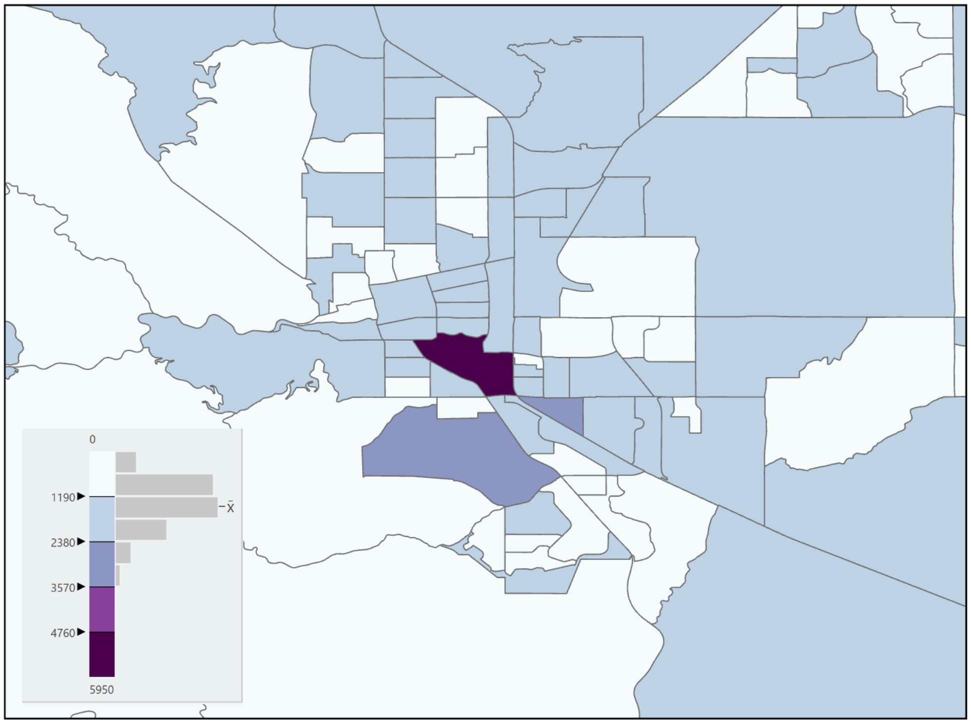


Figure 10.10. Equal Interval classification
Data Source: US Census Bureau

Equal Frequency/Quantile: The equal frequency classification method (Figure 10.11), also known as the quantile method, distributes observations equally among classes which means that each class will have the same number of observations. The advantages of this classification method are that it is easily calculated, applicable to ordinal data, and will not create a class with zero observations. The disadvantages of this method are that it does not consider the distribution of data, it can create gaps in the legend, and similar data values may not be grouped in the same class. Additionally, if the number of observations does not divide equally into the number of classes some classes will have more observations than others [3].

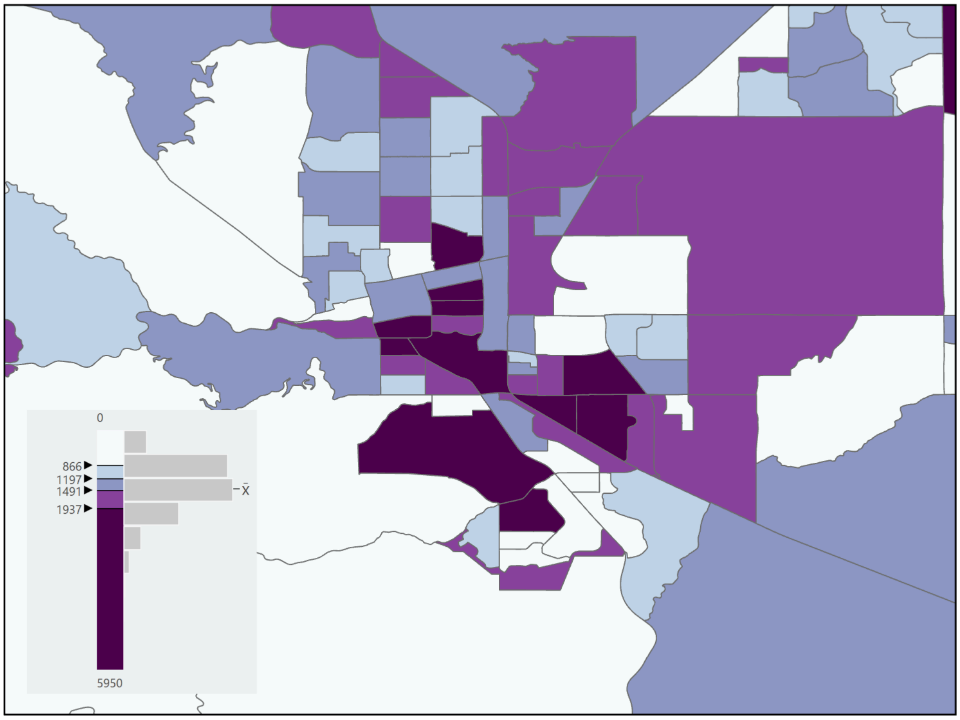


Figure 10.11. Equal Frequency
Data Source: US Census Bureau

Arithmetic and Geometric Intervals: The arithmetic and geometric intervals classification methods create class boundaries that change systematically with a mathematical progression (Figure 10.12). This classification method is useful when the range of observations is large and the observations follow some sort of mathematical progression that can be followed with the classes. The advantages are that it is good for data with large ranges and the breakpoints are determined by the rate of change in the data set. The disadvantage is that it is not appropriate for data with a small range or with linear trends [3].

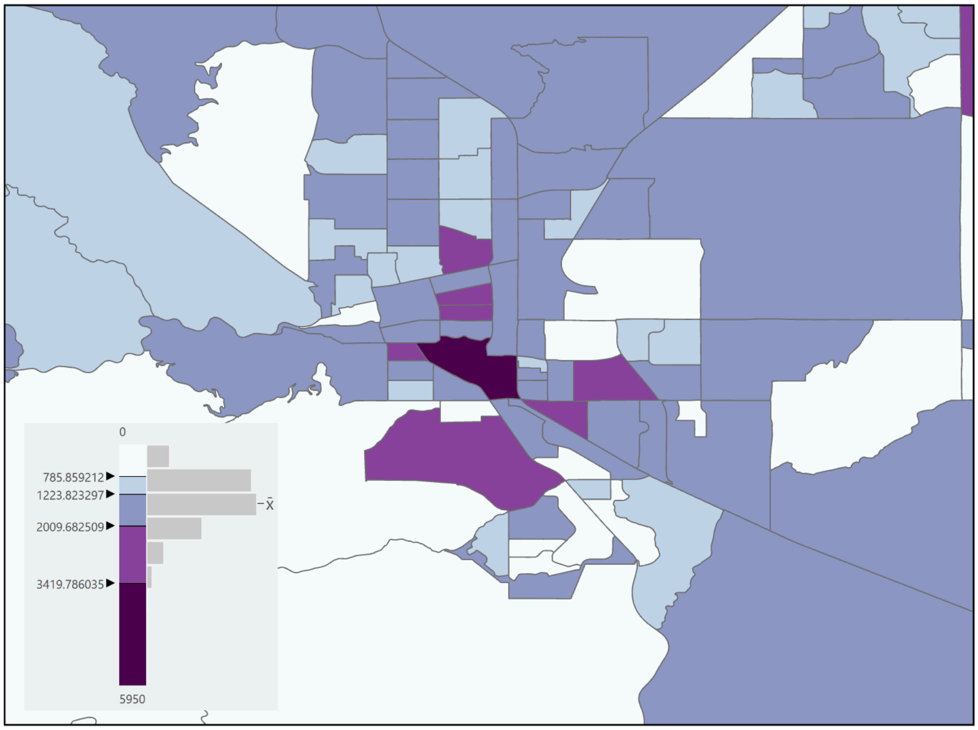


Figure 10.12. Geometric classification.
Data Source: US Census Bureau

Manual Classification: The map maker can define a classification method of their choosing which is known as manual classification. While the advantage is that the map maker has full control over the classification, the process can be time-consuming, the logic may not be apparent to the map reader, nor may the distribution of the data have been considered [1].

10.4 Patterns of Error

As classification is a type of generalization and with all types of generalization error is introduced, we can examine the “fit” and error introduced by a chosen classification scheme. The Goodness of Variance Fit (GVF), developed by George F. Jenks, provides one way in which we can examine the errors introduced, but also can guide the cartographer in choosing an appropriate number of classes for a particular scheme.

Goodness of Variance Fit (GVF): Determining the GVF is an iterative process. Calculations must be repeated using different classing methods for a given dataset to determine which has the smallest “in-class” variance, meaning that the data values within the classes are fairly homogeneous and “fit” the class limits.

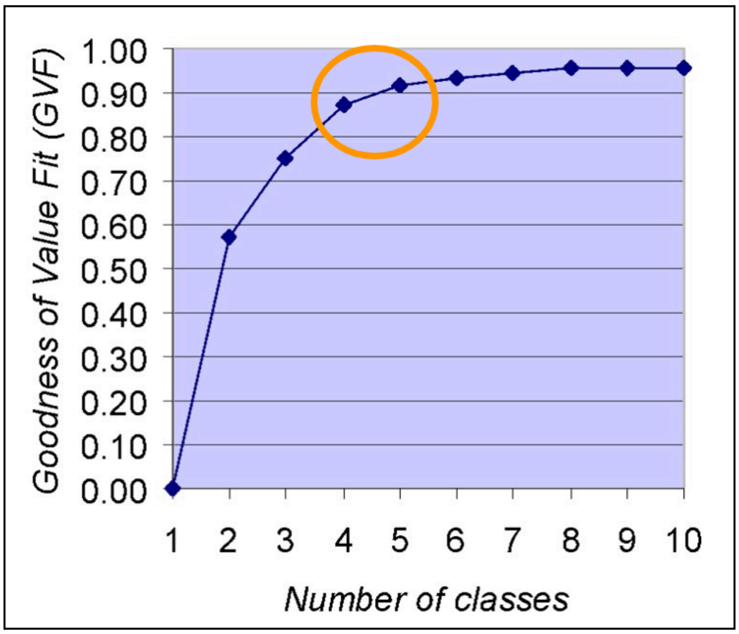
To determine the GVF, the following steps are performed:

1. Calculate the sum of squared deviations between classes (SDBC).
2. Calculate the sum of squared deviations from the array mean (SDAM).

3. Subtract the SDBC from the SDAM (SDAM-SDBC). This equals the sum of the squared deviations from the class means (SDCM).

Once these calculations have been run on a particular scheme, the map maker then chooses a different classification scheme and runs the calculations again. New SDCMs are then calculated, and the process is repeated until the sum of the within-class deviations reaches a minimal value. The GVF of a particular classification scheme is 1.0 – SDCM / SDAM. The closer the GVF is to one, the better the scheme fits the dataset; the closer to zero, the worse the fit [1][4].

The GVF can also be used to determine the appropriate number of classes to use for a particular dataset. By computing the GVF for different numbers of classes (e.g., 5 classes vs 7 classes) you can look for the point of "leveling off" of the GVF. Beyond that point, adding classes will notreduce the error enough to warrant the extra detail (Figure 10.13) [1].



# classes	GVF	relative % rise
1	0.000	----
2	0.575	58%
3	0.753	31%
4	0.874	16%
5	0.917	5%
6	0.934	2%
7	0.944	1%
8	0.953	1%
9	0.956	0%
10	0.958	0%

Figure 10.13. GVF for determining the number of classes.
Credit: Image by Barbara Buttenfield, University of Colorado Boulder, used with permission

10.5 Further Considerations for Data Classification

Encompass the Full Range of Data: The classification method used should always encompass the full range of data. Do not exclude data to meet a particular agenda. Any "no data" or null values should still be visualized but not as part of the classification scheme. Instead, the "no data" values should be visualized differently than the other classes and those features should be explicitly labeled in the legend. When looking for "no data" values in a data set you should read the metadata for the data set to see how they represent the no data values. Common no data values are: no data, -99, -9999, null, ND, NA [3].

No Overlapping or Vacant Classes: When classifying the data there should never be overlapping or vacant classes. For example, a class that holds 17 to 30 and values 30-45 could be changed to hold the values 17 to 29, for the first instance. If the values include decimal places, those should be used in the class breaks and indicated in the legend. Furthermore, there should never be vacant classes, which means a class that has no observations (e.g., no maps features are represented by that class). The only exception is if you're making a series of maps covering a wide time range and you want to use a single legend for all of the maps [3].

Choose an Appropriate Color Scheme: You should choose colors wisely when applying them to classes. Two categories of color choices are appropriate for classifying quantitative data: sequential and diverging (see Chapter 6).

Modifiable Areal Unit Problem (MAUP): The modifiable areal unit problem (MAUP) is a source of statistical bias that is common in spatially aggregated data, or data grouped into regions where only summary statistics are produced within each district. In other words, individual data points are summarized by a potentially arbitrary division, at least relative to the phenomena being studied. MAUP is particularly problematic in spatial analysis and choropleth maps, in which aggregate spatial data is commonly used..

The two types of biases/errors examined with MAUP are scale and zonal effects. The scale effect occurs when different levels of aggregation are used, such as counties versus census tracts. So while the same data values are used, each successively smaller unit changes the statistical pattern observed. The zonal effect is about the shape of the spatial divisions. For this type of error, each change to the division shape creates potentially significantly different results. For example, points aggregated at a county level versus zip codes will yield different results. For additional information on MAUP, go the the GIS&T Book of Knowledge https://gistbok.ucgis.org/bok-topics/problems-scale-and-zoning [opens in new tab]. It discusses the scale and zonal problems that exist in GIS and mapping, including ways in which to address the issues and it provides visual examples.

Additional Resources

Normalizing data: https://www.e-education.psu.edu/geog486/node/608 [opens in new tab]
Statistical Mapping:
https://gistbok.ucgis.org/bok-topics/statistical-mapping-enumeration-normalization-classification [opens in new tab]

References - materials are adapted from the following sources:

[1] GEOG 3053 Cartographic Visualization by Barbara Buttenfield, University of Colorado Boulder, used with permission.

[2] GEOG 486 Cartography and Visualization by Cary Anderson, Pennsylvania State University, under a CC BY-NC-SA 4.0 license.

[3] Introduction to Cartography by Ulrike Ingram under a CC BY 4.0 license

[4] Wikipedia, "Jenks natural breaks optimization", under a CC BY-SA 4.0 license

Chapter 11
Typography

In this chapter, you will be introduced to the concepts of typography relative to map making including the elements that make up typography and the effect it has on understanding a map. Type, characteristics of type, the various uses of type on maps, and how to place labels are included in this chapter. Finally, guidelines for the placement of labels based on the feature type (point, line, or area) are provided [5].

11.0: Introduction to Typography

Typography is the process of designing and placing a type on a map. Type is placed on maps to inform the users of things like the purpose of the map, the source of information, and the attribute values of the features displayed on the map. When the type is well placed it will greatly improve the usefulness of a map, along with its professionalism [5].

Placing type on a map can be a very time-consuming process as labeling routines within GIS software packages rarely do an adequate job of placing type on a map automatically for every label. Therefore it is typically up to the cartographer to manually adjust or place type on a map. Unfortunately, placing type on a map is often treated as an afterthought when it is actually one of the most important map elements for the map user [5].

Defining Type: Type is generally organized by typeface/font, style, size, and family. A typeface or font is a set of all the alpha, numeric, and special characters and the type family, style, and size associated with a particular font [5]. For an example of a character map that shows all of the characters in the Arial font visit Microsoft's Arial font family page https://learn.microsoft.com/en-us/typography/font-list/arial [opens in new tab].

Functions of Type: Type most often serves to name places or cartographic elements. It is a "functional symbol" primarily, and secondarily an aesthetic element meaning that it carries cartographic meaning and is not simply "window dressing". The four primary functions of type on a map are to label features, explain map elements, establish hierarchy or size of elements, and direct the reader's attention to particular features [3].

Thinking back to the visual variables discussed in Chapter 4, we can relate type to those visual variables. Specifically, size relates to importance, shape to the font style, texture to extent (character spacing), and position to where the label is placed relative to the feature [3].

Typefaces / Font: Typeface, commonly referred to as font, is the name associated with a particular set of design characteristics that are unique to that typeface. Typefaces include variations in x-height (size), weight (e.g., light, bold), form (e.g., italic), and letter spacing (e.g., condensed, monospaced). Commonly used fonts are Arial, Times New Roman, Garamond, Helvetica, and Tahoma. For a periodic table of commonly used fonts, including their style (e.g., san serif) and the year it was designed, visit the Periodic Table of Typefaces https://www.behance.net/gallery/193759/Periodic-Table-of-Typefaces [opens in new tab].

Font Style: The font style is a design variation of a typeface. One of the most common font styles is Roman, also called Oldstyle or Old Face, which has serifs. Serifs are small finishing strokes at the end of the letters. Serifs make the type easier to read by guiding the eye from letter to letter and word to word. The Roman style has a small x-height which means that the standard height of lowercase characters is generally short [5].

The sans serif style began to be heavily used in the 19th and 20th centuries. There is no variation in stroke thickness, no serifs, and a larger x-height than the Roman style. Modern font styles are similar to Roman as they also have serifs, but the serifs are squared-off rather than rounded. The decorative font family began to be used in the 19th and 20th centuries and is easily recognizable as heavily ornamented characters [5]; decorative fonts are not used to label map features but may be used in titles and logos [3]. The script font family began to be used in the 19th and 20th centuries and resembles handwriting; this style has no use in cartography. [5]. For examples of serif, san serif, and decorative fonts see Figure 11.0 and the Purdue OWL website https://owl.purdue.edu/owl/general_writing/visual_rhetoric/using_fonts_with_purpose/font_features.html [opens in new tab].

Serif Font Decorative
San Serif *Script*

Figure 11.0 Different font styles.

Type Form and Weight: Form refers to the presentation of the font in both its weight (Figure 11.1) and whether or not it is upright (i.e., italics). Using Italics slants the letters to the right and should be used to label water features whose label should also be in the color blue. Italics are also used to identify data sources and to identify the scientific name of a species. Thin, light, medium, bold, etc. refer to the weight of each letter and can be used to indicate hierarchy in feature labeling [5].

Helvetica Neue 25 Ultra Light
Helvetica Neue 35 Thin
Helvetica Neue 45 Light
Helvetica Neue 55 Roman
Helvetica Neue 65 Medium
Helvetica Neue 75 Bold
Helvetica Neue 85 Heavy
Helvetica Neue 95 Black

Figure 11.1. Type Weights
Credit: Helvetica Neue typeface weights, *Wikipedia*, public domain

Type Size: Type size is the scaling of a font which is typically measured in points with one point being equal to 1/72 of an inch [4] or 0.353 mm (72 pts = 1 inch). However, font size varies by the font type and the differences are demonstrated by the "x-height", literally the height of the letter x. The x-height varies by font type even when using the same font size (Figure 11.2).

Figure 11.2. A demonstration of x-height for the Adobe Garamond typeface.
Credit: Typography Line Terms, *Wikipedia*, Max Naylor, public domain

When considering type size on a map you should choose a reasonable upper and lower limit. Generally, six-point font is the smallest type size that is still easily readable by your audience, but a type size this small should be used sparingly. Type size typically corresponds with the relative size of features with larger or more important features using a larger type size, and less important smaller features a smaller type size [5]. When using different font sizes, note that differences of less than 15% in point size are typically not distinguished by the map reader so larger differences should be used [3].

Font/Letter Spacing: All fonts are designed so that each letter either occupies the same amount of space, regardless of the letter's width (monospaced), or takes up only the space needed for each character (proportional) (Figure 11.3) [5].

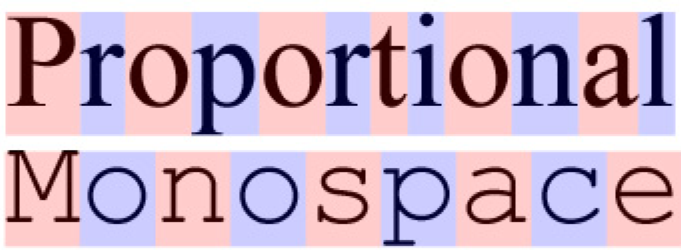


Figure 11.3. Examples of proportional and monospaced fonts.
Credit: Proportional vs monospace, *Wikipedia*, Garethlwalt, CC BY 3.0

Key Points for Using Fonts on Mapping: A map usually only encompasses two fonts, four maximum, with one serif and one sans serif font; never use only one font on the entire map. All of the fonts on a map should have similar "personalities" (e.g., modern vs Roman) and reflect the tone of the map's purpose [5].

Fonts, like any other software elements, are licensed to software packages, therefore, be cautious of using a font that may not be widely available in all software systems. Even fonts with the same name can still vary between software companies [3].

11.1 Type Characteristics

Case: Case refers to whether characters are displayed in UPPERCASE or lowercase form. Lowercase letters have been shown to be easier to read and distinguished from each other better because they have more detail in the lower half of the letter [5].

Letter Spacing: Letter spacing (Figure 11.4) is done by the mapmaker to add or reduce the distance between characters in a word. This is done in addition to what is already built into the font itself [5]. Extra letter spacing is used on maps when labeling large areas (oceans, canyons, mountain ranges, etc.) and is typically done so in all uppercase letters. It is also useful for improving the appearance of curved type such as labeling a river. Some GIS software programs allow the mapmaker to set the spacing between characters as a percentage of the text element point size (e.g., 200% means 2 letters will fit in the space between each letter) [3]. If you decide to modify the letter spacing it should be consistent within blocks of type and among types that represent similar phenomena [5].

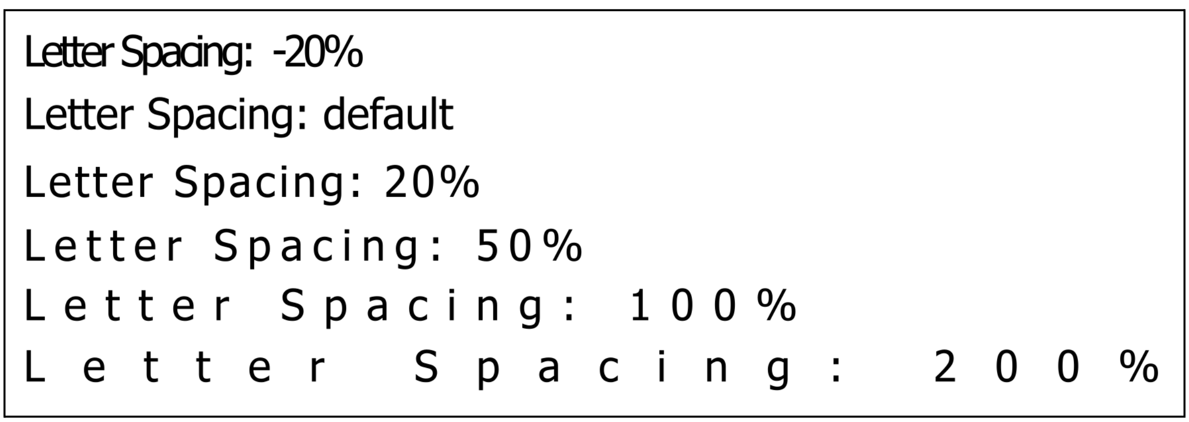


Figure 11.4. Letter spacing implemented using ArcGIS Pro.

Leading: Leading refers to the vertical space between lines of text; spacing between texts is measured in points (Figure 11.5) [5]. Large leading is used to label bigger areas whereas tight leading is used to label smaller areas and is also influenced by the map's feature spacing. Similar to letter spacing, this varies by font [3]. Leading should be consistent within blocks of type [5].

Figure 11.5. Text leading.

Overprinting: It is possible that in some cases your label will be overprinted on a map feature. When this happens it may obscure the map feature and make the map and label difficult to read. Fixes include adjusting the placement or employing effects such as halos, shadows, or callouts (see Section 11.2) [5].

11.2 Effects

Halo: Halos help make labels stand out from the surrounding areas, especially in busy portions of a map. Halos can be any color with white being the default in some GIS software programs (Figure 11.6). They can also be the color of the map background, especially when breaking features such as roads (Figure 11.7). Be careful not to make a halo too large because it may draw undue attention to the label [3].

Figure 11.6. White halo.
Credit: *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

Figure 11.7. Halo matching the background color.
Credit: *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

Callout/Leader Line: If a label is larger than the area it is supposed to be labeling or is in a congested area, callouts or leader lines can be used (Figures 11.8 and 11.9). These effects allow the label to be placed away from the feature directing the user's attention to the fact that the label belongs to a feature further away [5]. However, the first option is to consider placing the label differently, as callouts and leader lines should be used sparingly and only if a better placement is not an option. The font size of the label should not be adjusted to fit the label within a feature boundary as that can imply that the feature is of a different class or hierarchy [3].

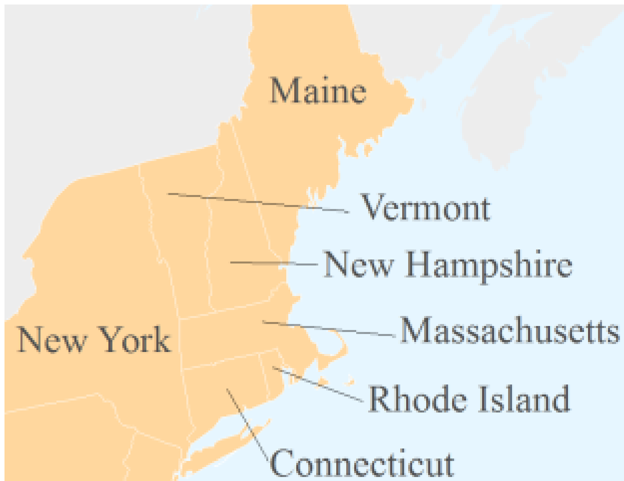


Figure 11.8. Leader lines are applied where labels will not fit within the feature boundaries.
Credit: *Introduction to Cartography*, Ulrike Ingram, CC BY 4.0

Figure 11.9. Balloon callout.

Shadows: Similar to halos, shadows, also called drop shadows, are used to increase legibility by providing contrast between the text and the map, especially in areas of a map where the background is very busy (Figure 11.10). However, drop shadows create the impression that the feature is raised off the page with the shadow of the letters showing to the right or lower right of the text. Any color or shadow thickness can be used, but as with halos, should be of a complimentary color and not too large as to draw unnecessary attention to the label.

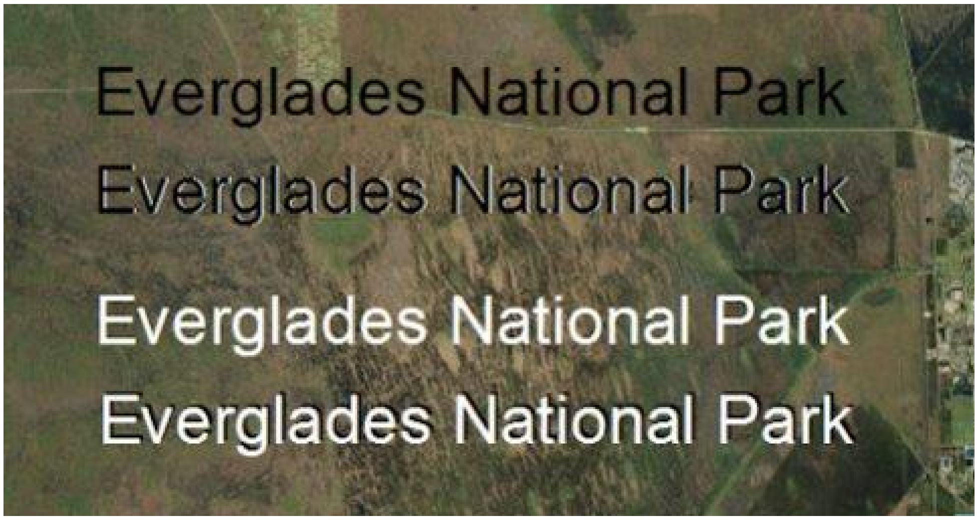


Figure 11.10. Labels without shadows (1st and 3rd) and with shadows (2nd and 4th).

11.3 Labeling Map Features

This section goes over key placement guidelines for labels relative to the type of feature being labeled - points, lines, and areas.

Point Label Placement: When placing point labels, two factors are of primary importance: (1) legibility, and (2) association. You do not want your reader to struggle to read your map labels, and it should always be clear to which point each label refers [4] (Figure 11.11). Other key guidelines for point placement include never increasing letter spacing, avoiding label overlap, ensuring all lettering of a label is placed upright (with no upside-down letters), and using callouts/leader lines sparingly [3].

Labels for points should be placed next to the point, horizontally, and with default letter spacing, but shifted up or down relative to the point as it is preferable over the label being directly aligned with the point. All labels and points should be placed at a consistent distance from each other; pay particular attention to this is manually placing or adjusting labels. For point features that are located along a coastline (e.g., lighthouse, coastal city), position labels fully in the water; do not cut the coastline with the label. The converse applies to inland features, the label should be fully on the land feature, not cutting the coastline (Figure 11.12). For labels that must cross map features, such as roads, rivers, or boundary lines, use halos, but minimize the frequency [5]. Labels for point features along or that are near linear features, should not be separated from the point. In other words, the label and point should be on the same side of the linear feature as is shown on the National Park Service map of campgrounds in Yosemite National Park (Figure 11.13) [3].

Figure 11.11. Poor label placement (left), ok label placement (middle), better label placement (right).
Credit: Image by Cary Anderson, Penn State University, [GEOG 486 Cartography and Visualization](). [CC BY-NC-SA 4.0]()

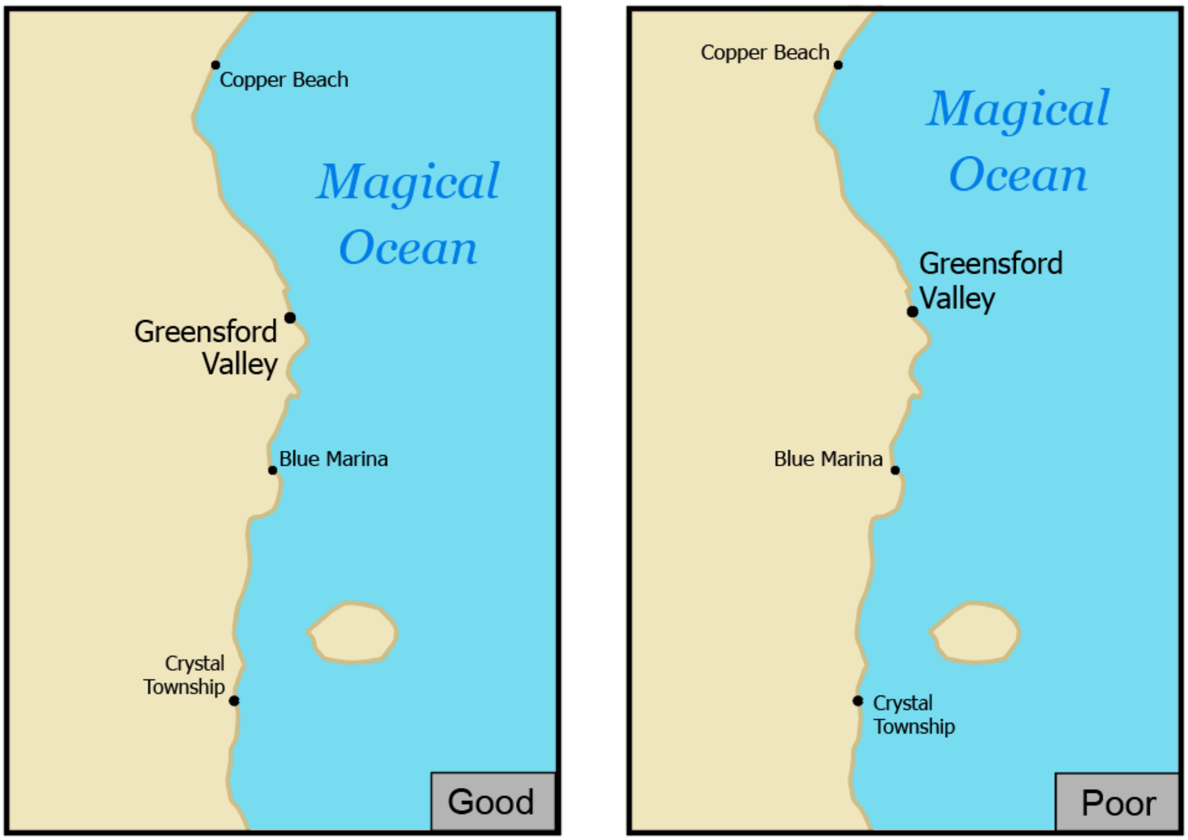


Figure 11.12. Good label placement (left), poor label placement (right). Coastal feature labels should be placed in the water, while land feature labels should be entirely on land.
Credit: Image by Cary Anderson, Penn State University, [GEOG 486 Cartography and Visualization](). [CC BY-NC-SA 4.0]()

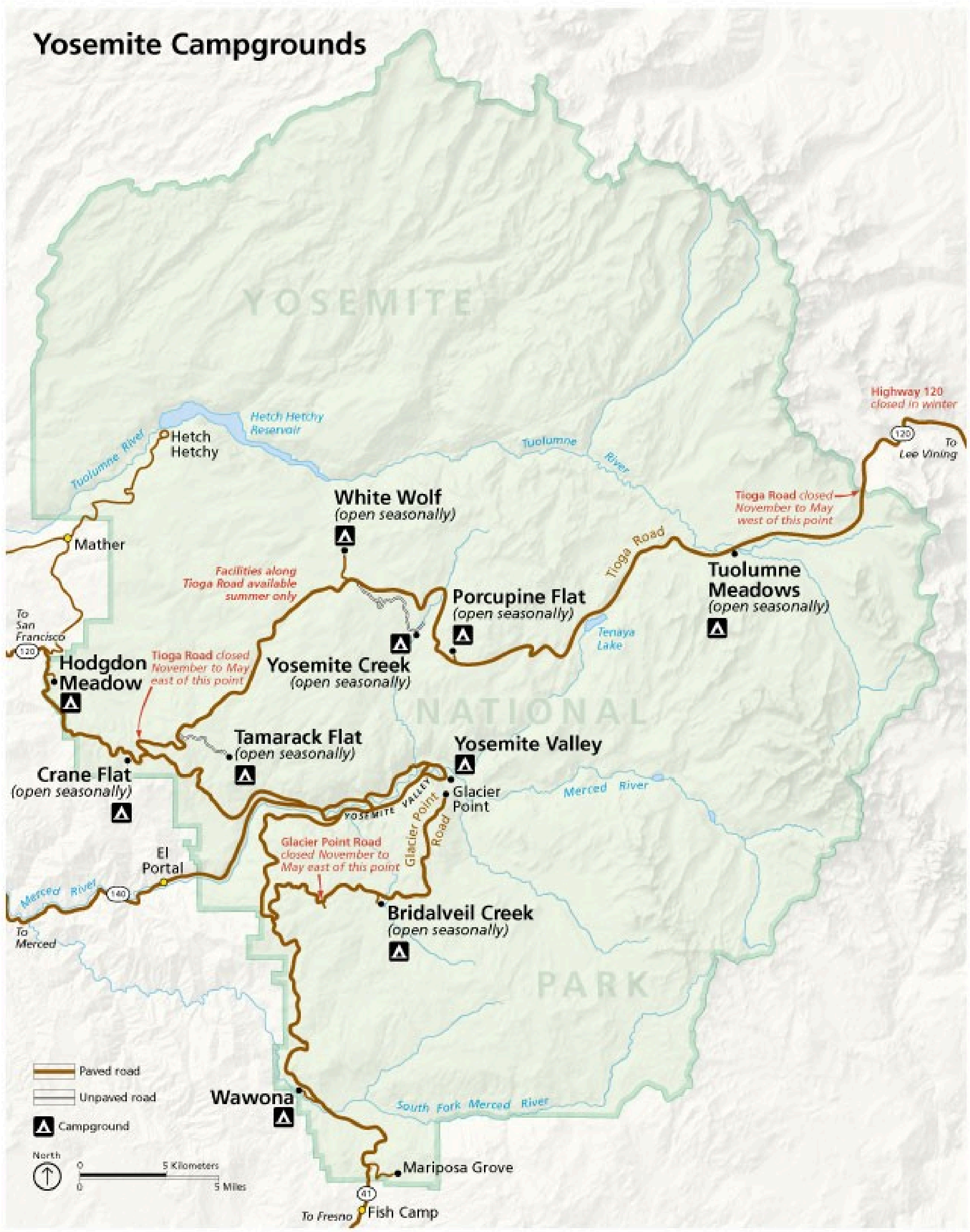


Figure 11.13 Yosemite National Park campground map showing effective point label placement.
Credit: NPS Map of Yosemite National Park, USGS, public domain

Line Label Placement: Line features, such as roads and hydrologic features, should have labels that follow the linear feature with a small gap between the label and the feature such that the gap is a consistent distance throughout the map. For choosing a font, use one with proportional letter spacing, not monospacing [3].

Avoid stretching out text (use default letter spacing), repeating the label along a line is a better solution. The labels should be placed above a feature as opposed to below it and the map maker should avoid upside down labels. To help achieve this, labels should be placed at the straightest and most horizontal portion of the feature. However, if the feature is relatively curved or curvy, the label can follow the linear feature, but not so closely that the letters start to collapse in on themselves. Also, do not allow part of the label to tilt upside down in any situation (Figure 11.14). Note that the figure used blue italic text with the directionality of the water feature indicated by the directional placement of the label. As with point label placement, breaking lines that interfere with labels need to be minimized and effects used as needed [3].

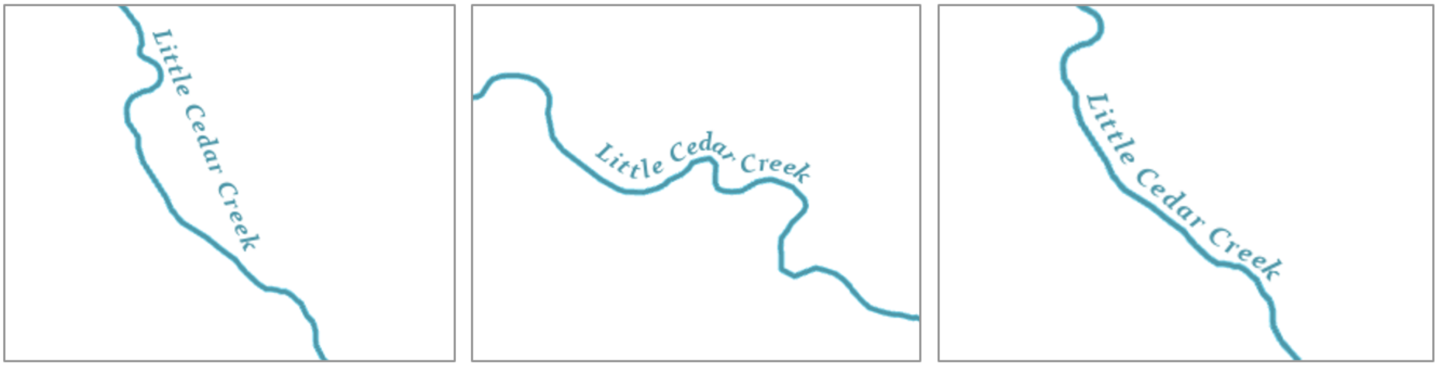


Figure 11.14. Poor river label placement (left and middle); Better label placement (right).
Credit: Image by Cary Anderson, Penn State University, [GEOG 486 Cartography and Visualization](). [CC BY-NC-SA 4.0]()

Area Label Placement: Areal features include oceans, forests, counties, states, and all features that are represented by polygons/areas. Area label placement differs from point and line label placement because they often use letter spacing which is best suited for uppercase lettering. However, area label placement should not be confused with the size of the area with the feature's visual hierarchy. In other words, a big area does not necessarily mean it's the most important feature or needs to be emphasized. Along with letter spacing, area labels, may use line spacing and position to suggest extent - a feature name is normally spread out to cover the area extent and curved to match the shape of the area (Figure 11.17) such as a mountain range. Do not simply tilt the label horizontally. As with other label placements, adjust the position so gaps fit across features and other labels. If the entire label does not fit within the area it is labeling, leader lines should be used (see Figure 11.8 above) [3].

Figure 11.16. Poor (left) and good (right) area label placement.
Credit: Image by Cary Anderson, Penn State University, [GEOG 486 Cartography and Visualization](). [CC BY-NC-SA 4.0]()

11.4 Visual Hierarchy in Label Placement

Establishing a visual hierarchy is an essential aspect of map design that helps map readers organize graphical information so that it can be understood more quickly and easily. When applied to labels and text, visual hierarchy allows the map reader to quickly and easily execute basic map reading tasks like categorizing, grouping, and locating important information on the map. Figure 11.18 shows a map without a label hierarchy. Without label hierarchy, map reading becomes more taxing because everything is assigned an equal level of importance [1]. Figure 11.19 is the same map with label hierarchy applied effectively. Note the use of uppercase lettering for countries and states, lowercase for cities, and blue italics for water features. Also, the state names utilize a different size and color than the country labels in order to differentiate them and establish relative order of importance on the map.

Figure 11.17. Map without label hierarchy.
Credit: Image from *Cartography Guide*, [Axis Maps](), [CC BY-NC-SA 4.0]()

Figure 11.18. Map with label hierarchy.
Credit: Image from *Cartography Guide*, Axis Maps, CC BY-NC-SA 4.0

All else equal, bigger and bolder styles tend to promote labels and text in the visual hierarchy. Capitalization and bold colors like black or red, can have a similar, promoting effect. Smaller point sizes and less letterspacing tend to demote labels in the visual hierarchy, as do more muted text colors, like light brown (Figure 11.19). Of course, the task becomes more complex as more labels are added and the number of necessary styles increases. Creating a good visual hierarchy will normally require some experimentation, revision, and refinement [1]. For examples of different ways in which visual hierarchy can be applied to labeling, including weight, leading, hue, and value see the GIS&T Book of Knowledge figure on visual variable use for typography https://gistbok.ucgis.org/sites/default/files/Figure%206%20visual%20variables.png [opens in new tab].

Additional Resources
Typography: https://gistbok.ucgis.org/bok-topics/typography [opens in new tab]

References - materials are adapted from the following sources:

[1] Cartography Guide by Axis Maps under a CC BY-NC-SA 4.0 license

[2] Essentials of Geographic Information Systems by Saylor Academy under a CC BY-NC-SA 3.0 license

[3] GEOG 3053 Cartographic Visualization by Barbara Buttenfield, University of Colorado Boulder, used with permission.

[4] GEOG 486 Cartography and Visualization by Cary Anderson, Penn State University, under a CC BY-NC-SA 4.0 license.

[5] Introduction to Cartography by Ulrike Ingram under a CC BY 4.0 license